DOCUMENT RESUME

ED 142 446 SE 022 915

AUTHOR Herbert, Martin; And Others
TITLE Evaluation Report 3-B-1. Second and Third Grade Test Data, Year 3.
INSTITUTION Central Midwestern Regional Educational Lab., Inc., St. Louis, Mo.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE [76]
NOTE 193p.; Extended Pilot Trial of the Comprehensive School Mathematics Program; For related documents, see SE 022 895, 908-916 and ED 101 993-ED 102 007; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Achievement; *Curriculum; Elementary Grades; *Elementary School Mathematics; *Evaluation; Instruction; Instructional Materials; Mathematics; Mathematics Education; *Program Evaluation
IDENTIFIERS *Comprehensive School Mathematics Program

ABSTRACT

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials in mathematics for grades K-6. Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. This report summarizes test data collected from second and third grade classes during 1975-76, the third year of the CSMP Extended Pilot Test. Included in this report is an introduction, descriptions of tests used, discussion of data, and Appendices. The appendices include data from the study and copies of most instruments. On standardized mathematics tests, CSMP classes generally did as well as or better than non-CSMP classes. On tests relating to CSMP content, CSMP classes scored significantly higher than non-CSMP classes. Results were similar to data from previous studies. (RH)

CSMP Evaluation Report Series

Evaluation Report 3-B-1

Second and Third Grade Test Data, Year 3

Martin Herbert
Knowles Dougherty
Ava Small

Developed by CEMREL, Inc., a private nonprofit corporation supported in part as an educational laboratory by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

Description of Evaluation Report Series

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials for grades K-6.

Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. The pilot trial is longitudinal in nature; students who began using CSMP materials in kindergarten or first grade in 1973-74, were able to use them in first and second grades respectively in 1974-75, and in second and third grades in 1975-76. Hence the adjective "extended".

The evaluation of the program in this extended pilot trial is intended to be reasonably comprehensive and to supply information desired by a wide variety of audiences. For that reason the reports in this series are reasonably non-technical and do not attempt to widely explore some of the related research issues. The list of reports from the first two years of the extended pilot trial is given on the next page. The most comprehensive of these are the following:

1-A-1: Overview, Design and Instrumentation
1-A-3: Final Summary Report, Year 1
2-A-1: Final Summary Report, Year 2
and 3-B-1: Summary of Second and Third Grade Test Data Year 3

Longitudinal Pilot Study of the
Comprehensive School Mathematics Program

Evaluation Report Series

Evaluation Report 1-A-1	Overview, Design and Instrumentation
Evaluation Report 1-A-2	External Review of CSMP Materials
Evaluation Report 1-A-3	Final Summary Report Year 1
Evaluation Report 1-B-1	Mid-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-2	End-of-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-3	End-of-Year Test Data: Standard First Grade Content
Evaluation Report 1-B-4	End-of-Year Test Data: CSMP Kindergarten Content
Evaluation Report 1-B-5	Test Data on Some General Cognitive Skills Related to CSMP Content
Evaluation Report 1-B-6	Summary Test Data: Detroit Schools
Evaluation Report 1-C-1	Teacher Training Report
Evaluation Report 1-C-2	Observations of CSMP First Grade Classes
Evaluation Report 1-C-3	Mid-Year Data from Teacher Questionnaires
Evaluation Report 1-C-4	End-of-Year Data from Teacher Questionnaires
Evaluation Report 1-C-5	Interviews with CSMP Kindergarten Teachers
Evaluation Report 1-C-6	Analysis of Teacher Logs
Evaluation Report 2-A-1	Final Summary Report Year 2
Evaluation Report 2-B-1	Second Grade Test Data
Evaluation Report 2-B-2	Readministration of First Grade Test Items
Evaluation Report 2-B-3	Student Interviews
Evaluation Report 2-C-1	Teacher Questionnaire Data
Evaluation Report 2-C-2	Teacher Interviews, Second Grade
Evaluation Report 2-C-3	Teacher Interviews, First Grade
Evaluation Report 3-B-1	Second and Third Grade Test Data Year 3
Evaluation Report 3-C-1	Teacher Questionnaire Data Year 3

Key to Indexing

1-C-2 Observations of CSMP First Grade Classes

"2" refers simply to the number within a given year and type of data

"C" refers to the type of data being reported
- A: Overview, summary and theoretical reports
- B: Student outcomes
- C: Non-test data

"1" refers to the year of the Pilot Study according to the following:

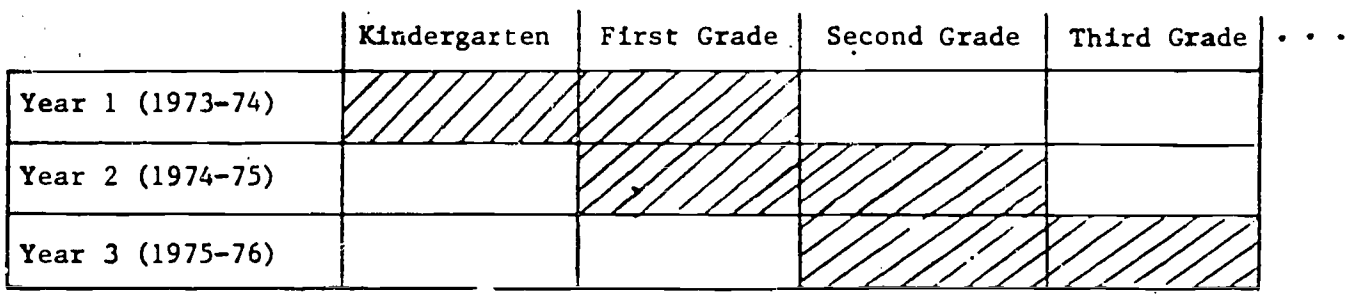


Table of Contents

Appendices

Introduction

This report summarizes test data collected from second and third grade classes during 1975-76, the third year of the CSMP Extended Pilot Test. This pilot test began in 1973-74 in kindergarten and first grades, with over 100 classes using CSMP materials. In succeeding years, many of these classes continued with CSMP in progressively higher grade levels as the curriculum materials became available, and many new classes joined the pilot test.

Much evaluation data was collected during the previous two years and is reported in the Evaluation Reports described on page iv. The major emphasis in this 1975-76 study was at the second grade level, where classes were using a revised version of second grade materials (CSMP Mathematics for the Upper Primary Grades Part I). Some data was also collected from third grade (CSMP Mathematics for the Upper Primary Grades Part II).

Three kinds of tests were used to investigate student learning:
a) Standardized Mathematics Test
b) CSMP-Specific Tests
c) Mathematics Applied to Novel Situations (MANS Test)

A description of the content and results from each of these sets of tests is given in each of the next three chapters respectively. The CSMP-Specific Tests were given only to CSMP classes, while the other two kinds of tests were given for comparison purposes to both CSMP and Non-CSMP classes. When making such comparisons, scores from a general ability test or a reading test were used as covariates. The use of such scores as covariates allowed scores from tests a) and c) above to be adjusted to take into account possible differences in ability and thus eliminate, or at least reduce the plausibility of, the explanation that differences in performances were a result only of differences in general ability.

Second Grade Testing

In addition to classes from four school districts in the St. Louis area (Normandy, Ferguson-Florissant, Ladue and St. Louis), test data was collected from several other school districts: New Hartford, N.Y.; Detroit, Mich.; Portland, Ma.; Clarksville/Montgomery County, Tenn.; Polk County, Ga.; Haralson Co., Ga.; and the Diocese of Marquette, Mich. Through what were called "Joint Research Studies", each of these districts cooperated in collecting some or all of the test data described above (limited by such things as availability of comparison classes, testing time, etc.). The school districts, through the coordinators, were responsible for selecting comparison classes (where available), scheduling the testing, and training the testers. Wherever possible, scores from regularly scheduled, district-wide tests were used as the standardized mathematics test and as covariates. The CSMP-Specific Tests and the MANS Test were developed by the CSMP Evaluation staff.

Thus a fairly complicated design emerged; in part a series studies from individual sites with somewhat different data collected from each, but with enough commonality that data could in some cases be combined across sites. One of the ways in which this was done was by transforming scores from the different covariates to percentile ranks (though it is recognized that percentile rank has some undesirable properties, so that different norming procedures must inevitably lead to some discrepancies between different covariates).

The testing plan for second grade classes is shown in Table 1, below. Site 01 was for classes in the St. Louis area; otherwise sites are not identified beyond the "site number" given in Table 1, and used in subsequent presentations of results.

Table 1

Testing Plan, Second Grade

Site	Comparison Classes?	Covariate	Standardized Math Test	Number of CSMP Classes/Non-CSMP Classes			Number of CSMP Classes, CSMP-Specific Tests		
				Standardized Math Test	MANS Test A	MANS Test B	Test A	Test B	Individual[1]
01	Yes	Kuhlmann Anderson	CTBS[2]	15/13	8/7	7/6	8	7	
12	Yes	SAT-Reading	SAT[3]	6/6	3/3	3/3	3	3	6
13	Yes	CPT-Reading	CPT[4]	6/6	3/3	3/3	3	3	6
25	Yes	Gates McGinitie	CTBS[2]	6/6	3/3	3/3	3	3	6
32	Yes	MAT-Reading[5]	CTBS[2]	1/1			1		
35[6]	Yes	Lorge-Thorndike	CTBS[2]	2/2	1/1	1/1	1	1	2
42	No						3	2	5
43	No						6	6	12

[1]Number of classes, from each of which 4 students were tested.
[2]Comprehensive Test of Basic Skills; Mathematics Computation, and Mathematics Concepts and Applications.
[3]Stanford Achievement Test; Mathematics Computation and Mathematics Concepts.
[4]Cooperative Primary Test, Mathematics
[5]Metropolitan Achievement Test
[6]Covariate data from this site not available in time to be included in analysis.

Third Grade Testing

With the exception of two non-local third grade classes, from whom no reliable covariate data was available, the testing was confined to three school districts in the St. Louis Metropolitan area. Altogether there were seven CSMP third grade classes locally and these classes were a part of the "lead" group of classes who began using CSMP in first grade in 1973-74.

Eight Non-CSMP classes were selected as comparison classes. These were either the "other" third grade class in the school or a class from an adjacent school. In the fall of 1975 both CSMP and Non-CSMP classes were administered the Kuhlmann-Anderson Test as a covariate. In the spring both groups of classes were administered the mathematics subtests of the Comprehensive Test of Basic Skills; and each CSMP class was administered one of two CSMP-Specific Tests.

The chapters in this report have been kept fairly short and non-technical. Much more detail can be found in the various appendices which make up the bulk of the report.

Standardized Math Tests

Second Grade

In order to measure students' knowledge of basic arithmetic skills, standardized tests were administered to CSMP and to Non-CSMP classes. In the local St. Louis area (denoted "site 01", but made up of classes from four school districts), the Mathematics test of the Comprehensive Test of Basic Skills (CTBS) was administered by specially trained CEMREL testers. In Outer Ring sites, the usual policy was to use already existing data from whatever standardized test was routinely administered in the spring by the individual school district. Although this made for good economy, it is impossible to lump these different sites together in one analysis, and the analysis was therefore done separately at each site. In the local area it was possible to also collect data at the item level.

The general procedure for each site was to compare class means from CSMP classes with those from Non-CSMP classes, using scores from some tests of general ability or reading to adjust for potential differences in ability between these two groups. For each site, the standardized math test used and the test used as a covariate for adjusting scores are given below.

Site 01:

Standardized Math Test: CTBS (raw score).
At Level C (Grades 1.6 - 2.9), the mathematics section is divided into two separately timed tests: Mathematics Computation (28 items) and Mathematics Concepts and Applications (25 items). The Computation section consists of 10 addition, 10 subtraction, and 8 multiplication problems. The Concepts and Applications section attempts to measure skills in basic operations: numbers, numeration, measurement, and fractions. The problems are read aloud to the students, who select their answers from pictured, numerical, or printed responses.

Covariate: Kuhlmann-Anderson (raw score) 10/75

Number of Classes: 15 CSMP; 13 Non-CSMP.

Site 12:

Standardized Math Test: Stanford Achievement Test (SAT) (percentile rank)

Covariate: Reading Test, SAT (percentile rank) (5/76)

Number of classes: 6 CSMP; 6 Non-CSMP

Site 13:

Standardized Math Test: Cooperative Primary Test (CPT) (raw score)

Covariate: Reading Test, CPT (raw score) (5/76)

Number of Classes: 6 CSMP; 6 Non-CSMP.

Site 25:

Standardized Math Test: CTBS (raw score)

Covariate: Gates McGinitie Reading Comprehension (raw score) (5/76)

Number of Classes: 6 CSMP; 6 Non-CSMP.

Second Grade Results

In order to compare test performance of CSMP vs. Non-CSMP classes, an analysis of covariance was performed at each site, using class means as the unit of analysis. The results are shown in Table 2, below, including - for each site - means across classes before and after adjustment for the covariate and the p-value of the difference between CSMP and Non-CSMP means. (The p-value is the probability of such observed differences occurring merely by chance - a low p-value indicates that chance alone is an unlikely explanation.)

Table 2

Summary Data from
Standardized Math Tests: Second Grade
(First Entry: CSMP Classes, Second Entry: Non-CSMP Classes)

Site (Number of CSMP, Non-CSMP Classes)	Covariate[1]	Computation			Concepts and Applications		
		Means	Adjusted Means	p value	Means	Adjusted Means	p value
Site 01 (15,13) Comprehensive Tests of Basic Skills (Raw Scores)	48.8 46.8	21.0 18.6	20.6 19.1	.11	18.7 17.8	18.3 18.2	.86
Site 25 (6,6) CTBS (Raw Scores)	21.5 20.7	20.9 18.2	20.5 18.6	.11	39.4 35.3	18.1 17.5	.48
Site 35 (1,1) CTBS (Raw Scores)	54.8 76.7	24.5 23.8	too few classes		22.4 21.6	too few classes	
Site 12 (6,6) Stanford Achievement Test (Percentile Ranks)	57.7 42.0	54.9 44.4	47.0 52.3	.41	53.3 37.9	43.5 47.7	.23
Site 13 (6,6)[2] Cooperative Primary Test (Raw Scores)	32.8 35.4				35.7 36.3	36.2 35.7	.74

[1]Covariate: Site 1: Raw score, Kuhlman Anderson Test (10/75)
Site 25: Raw score, Gates McGinitie Reading Comprehension (5/76)
Site 35: Percentile Rank, Reading Test, Metropolitan Achievement Test (10/75)
Site 12: Percentile Rank, Reading Test, Stanford Achievement Test (5/76)
Site 13: Raw score, Reading Test, Cooperative Primary Test (5/76)

[2]At Site 13, the Cooperative Primary Test has only a total score in mathematics, which data has been put in the "Concepts and Applications" columns.

It can be seen that, while the differences were usually in favor of CSMP classes, in no case were these differences statistically significant. The closest any differences came to significance (usually taken as <.05) was on the computation section at Sites 01 and 25 where the p-value was .11. Appendices A (for Site 01) and B (other sites) provide information summarized according to class means. For each of the tests, the set of class means are plotted against the mean ability or reading score (transformed to percentile rank) of each class. By inspecting these graphs it is possible to determine and compare the extent to which various classes do about as one would expect based on their mean ability scores.

Where item analysis data was available (for the CTBS at Site 01) a compilation was made of all items in which the difference in percent correct between CSMP students and Non-CSMP students was more than 10 points. These items are shown below; in each case the difference was in favor of CSMP students. Complete item analysis data for this test is given in Appendix A.

Addition

$$\begin{array}{r} 154 \\ +\ 128 \end{array} \qquad \begin{array}{r} 27 \\ +\ .8 \end{array} \qquad \begin{array}{r} 13 \\ 6 \\ 52 \\ +\ 27 \end{array} \qquad \begin{array}{r} 1.14 \\ +\ .63 \end{array} \qquad 26 + 21 =$$

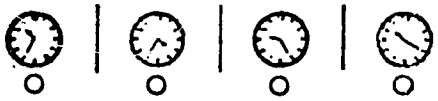
Subtraction

$$\begin{array}{r} 768 \\ -\ 427 \end{array} \qquad \begin{array}{r} 64 \\ -\ 9 \end{array}$$

Multiplication 5x2= 1x4= 2x3= 3x5= 3x2= 4x4= 8x5=

Concepts & Applications

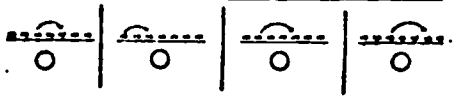
12. Find the clock that shows ten thirty-five.



15. Find the numeral that shows one-half of four.

13. Find the numeral that is one-half of six.

23. Find the number line that shows forty-five plus three.



Third Grade

The Comprehensive Test of Basic Skills was administered to the seven local St. Louis area third grade classes and to eight Non-CSMP comparison classes. At Level 1 of the CTBS (Grades 2.5 - 4.9), the mathematics section is divided into two separately timed tests: Mathematics Computation (48 items) and Mathematics Concepts and Applications (50 items).

Mathematics Computation: The 48 items in this test consist of 12 problems each in addition, subtraction, multiplication and division.

Mathematics Concepts & Applications: There are 50 items in all. The 25 "concepts" items purport to measure the student's ability to convert concepts expressed in one numerical, verbal or graphic form to another form, and to comprehend numerical concepts and their interrelationships. The content categories for these items include number systems, measurement, set theory, geometry and numeration. The 25 "applications" items purport to measure the student's ability to select and carry out problem-solving operations. The content categories include set theory, algebra, measurement, and reasoning.

Third Grade Results

An Analysis of Covariance was performed using class mean scores as the unit of analysis. The mean raw score from the Kuhlmann-Anderson test, administered 10/75, was used as the covariate. The results are given in Table 3 below.

Table 3

Summary Data from the Comprehensive Tests of Basic Skills: Third Grade, Local Classes
(First Entry: CSMP Classes (n=7), Second Entry: Non-CSMP Classes (n=8))

Test	Mean Scores	Adjusted Means	p value
Covariate[1]	49.8 53.2		
Addition Items	10.0 9.3	10.2 9.1	.09
Subtraction Items	9.3 8.9	9.5 8.6	.18
Multiplication Items	9.3 9.3	9.5 9.1	.63
Division Items	7.7 6.7	8.1 6.4	.25
Computation Test (total of above)	36.3 34.2	37.2 33.3	.03
Concepts and Applications	33.5 31.5	34.5 30.4	.02
Total Math Score	69.8 65.3	71.7 63.3	.01

[1]Mean Raw Score, Kuhlmann Anderson Test, administered 10/75.

Each of the four sub-parts of the Computation test has been analyzed separately, though none was significant at the .05 level. However, the differences were significant for the Computation test, for the Concepts and Applications test and for the total CTBS mathematics score; and all differences, whether significant or not, favored CSMP classes.

Appendix C shows the distribution of class means. Also given in that appendix is item analysis data for all the math items of the CTBS. Given below are those items for which the difference in percent correct between CSMP and Non-CSMP students was more than 10 points. (In each case the difference was in favor of CSMP students.)

Addition: 72+73 | 32 / 10 / + 44 | 26+13+14

Subtraction: 149 − 87

Multiplication: 25 × 4

Division: $3\overline{)9}$ $2\overline{)8}$ $5\overline{)25}$ $4\overline{)124}$ $24\overline{)48}$

Concepts & Applications:

23 Which of the following makes this number sentence true?

$3 \times (2 \times 4) = (3 \times 2) \times \square$

Ⓐ 4 Ⓑ 6 Ⓒ 8 Ⓓ 24

37 What is the missing number in 3, 5, 7, ____, 11?

Ⓐ 1 Ⓑ 8 Ⓒ 9 Ⓓ 10

38 What goes in the box to make this number sentence true?

$\square + 7 = 4 + 8$

Ⓐ 5 Ⓑ 6 Ⓒ 12 Ⓓ 19

39 What goes in the box to make this number sentence true?

$5 + 4 = 10 - \square$

Ⓐ 0 Ⓑ 1 Ⓒ 9 Ⓓ 10

CSMP-Specific Tests

Second Grade

Two group-administered tests, CSMP Test A and CSMP Test B, were developed to investigate student skills in working with the second grade CSMP content. The tests specifically excluded most of the traditional arithmetic skills and concepts of second grade since these were covered in the standardized mathematics tests described in the last chapter. Instead these tests focused on the special pedagogical "languages" of CSMP: arrow diagrams, the Minicomputer and string pictures.

The format for these tests was that of a CSMP workbook, an eight page booklet printed in color on newsprint. The printed directions were as similar as possible to those given in CSMP workbooks. The actual test pages were typical of the kinds of pages done by CSMP students on a regular basis; no new kinds of items were added nor were there novel applications or extensions of previous workbooks.

Administration of these tests was on a sampling basis; each CSMP class took either Test A or Test B. (The testing plan, across sites, is shown in Table 1, page 2.) These tests were not parallel tests; the sampling was done for the purpose of reducing the amount of time required for testing in a given classroom. Specially trained testers administered the test in a fairly straightforward way. After the tester reviewed one or two of the pages with the class, the students simply went ahead and worked right through it on their own, just as they would an ordinary CSMP workbook. Students raised their hands when they had questions, at which time the tester would explain the task individually to the student (though not, of course, how to do it). The test was essentially untimed. Most students took about 20 minutes to complete it; but up to 35 minutes was allowed if necessary.

For purposes of analysis, responses from pages dealing with similar areas of content were combined, though these pages might not be adjacent on the actual test. In this way the subtests named below were created. Also given below are some summary data.

	Number Items	Mean Percent Correct	Correlation with Kuhlmann-Anderson	Reliability (KR20)
Test A: Arrow Diagrams	12	72	.67	.87
Minicomputer	10	66	.64	.86
Integers	10	50	.39	.76
Test B: Arrow Diagrams	17	72	.72	.91
Minicomputer	10	75	.62	.90
String Pictures	4	65	.51	.63
Computation	7	69	.74	.80

Except for the subtest "Integers", the mean percent correct across items for the various subtests was between 65 and 75 and the correlation with the Kuhlmann Anderson Test was moderately high. The KR20 reliability of the subtests was also high with the exception of the very short (4 items) subtest "String Pictures".

An evaluation of the quality of student responses to this test is difficult for two reasons. First, because arrow diagrams and Minicomputers are not used in other math curricula, the tests were administered only to CSMP students, so that CSMP - Non-CSMP comparisons are not possible. Second, the test was not a mastery test and there are no specified standards or expectancies of the tasks and levels of success of which CSMP students should be capable. Indeed, because of the spiral nature of the CSMP curriculum, the concept of mastery is a much more difficult one to deal with. Summarizing statistics (such as mean percent correct) provide in themselves, very little useful information to the reader in the absence of the actual tests and some knowledge of the curriculum.

Part of Appendix D presents information summarized across students. For each of the seven subtests of Tests A and B, reduced copies of the actual test pages are given, together with a) percent correct for each item, and b) summary statistics (means, frequency distributions, correlations) across all students who took the test. Also given is similar information when students are grouped according to ability level. Thus, it is possible to determine on various items and subtests, what kinds of differences in performance there were between high and low ability students. Appendix D contains a graph for each subtest which shows the distribution of class means as a function of class ability scores.

In addition to the group administered CSMP Tests A and B, an individually administered test was constructed and administered to four representative students in each of 37 classes. These classes were all outside the St. Louis area and the tests, requiring about 15-20 minutes per student, were administered by special testers trained by the CSMP coordinators in the various sites.

The purpose of this test was to investigate, in situations not amenable to paper and pencil, students' skill in dealing with various tasks of the CSMP curriculum. Hence many of the test items were not typical of material found in the regular workbooks. They were probably more oriented to teacher-led lessons than to workbooks; several being rather challenging. There is simply no way to briefly summarize the results except to refer the reader to the actual test items and item statistics in Appendix E.

Third Grade

Two group-administered tests, CSMP Test A and CSMP Test B were developed to investigate student skills in working with the third grade CSMP content. These used the "languages" of the curriculum unique to CSMP. The various test pages were not similar in all cases to previously completed workbook pages; occasionally they were slight extensions of such pages.

These tests were also administered by special testers. Each of the seven third grade CSMP classes in the local St. Louis area took either Test A or Test B. Thus, the item analysis data, summarized in Appendix F (which also reproduces the test items), is based on at most four classes and no attempt has been made to further analyze this data by student ability or by mean class scores.

MANS Tests
(Mathematics Applied to Novel Situations)

Background

Since the beginning of the CSMP Extended Pilot Test, considerable time and effort has been devoted in the evaluation effort to the development and use of tests of student outcomes which go beyond the "basics" of the standardized tests yet do not rely on the specific "languages" of the CSMP curriculum. The rationale behind this effort was as follows: CSMP is interested in developing a math curriculum for elementary schools which would not only leave the "graduates" with the skills and concepts required in the traditional arithmetic curriculum, but would go beyond that in giving the student a foundation of understanding for mathematics itself.

Thus it was necessary to devise test items which would not only measure a student's grasp of higher-order ideas in mathematics, but would also do so in such a way as to give no superficial advantage to a student who had been exposed to mathematics in a particular way. Hence the items had to be at once (a) mathematical rather than merely arithmetic, (b) contextually novel, and (c) free of language peculiar to CSMP or any other specific elementary math curriculum.

During the first two years of the EPT, these items were grouped to form what were called "comparison tests." One can review the items developed and the results of their use in testing during those two years by consulting Evaluation Reports 1-B-5 and 2-B-1. During the third year of the EPT, the name for these tests was changed from "comparison" to MANS in order to be more descriptively accurate.

One bias should be pointed out here. Items were selected which were thought to be related to basic mathematical ideas behind the CSMP curriculum. Since the ideas are basic to mathematics, it might also be the case that other curricula would be aimed at giving students an understanding of the same ideas. Nevertheless, no attempt was made to select the items completely at random with respect to the basic mathematical ideas involved.

Description of MANS

The MANS test consists of ten subtests, divided into two booklets of five subtests each. The two booklets are not parallel in any sense, but have the same physical appearance and are administered in roughly the same way. Each booklet consists of eight 8½x11 newsprint pages. Each subtest is contained on one or two pages. Because of testing time considerations any participating class of students either took booklet A or booklet B but not both.

Specially trained testers administered the test in a fairly straightforward way. For each subtest the administrator gave directions orally and usually worked through the first item with the students. Each test was separately timed, so that all students began each new test together. Except for test A4 which was intentionally speeded, the time limits, based on pretesting, allowed most students to attempt each item. Each of

the booklets usually took about 40 minutes each to administer. Students who completed the subtests quickly were told to check their work and then to draw a picture on the front cover of the booklet if they so desired.

A sample test item for each MANS subtest is given below, together with a portion of the directions for that subtest. The full set of test items together with more complete directions is given in Appendix G.

Booklet A

Subtest 1: Sequences The crucial directions were, "The numbers are in a certain order...figure out what the pattern is...put the right number in the box."

Example: 16, 13, 10, ☐, 4, 1

Subtest 2: Equation Fluency The crucial directions were, "...write as many true number sentences as you can, using only these signs and numbers."

Example: = + - x 1 2 3

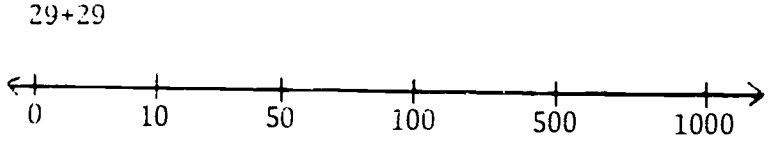
Subtest 3: Functions The crucial directions were, "A teacher was playing a game with the class...a student gave her 2 (the first number of each pair), she did something to it and got 4 (the second number) ...figure out what it was that the teacher was doing to the numbers, and then put the right number in the empty box."

Example:

2	4
5	10
7	14
10	

Subtest 4: Number Line Estimation The crucial directions were, "This is a funny looking number line isn't it? ...there are a whole lot of problems (13) and you won't have much time (2½ min.) to do them. You should not try to calculate the exact answer; just decide quickly where the answer would probably go on the number line."

Example:



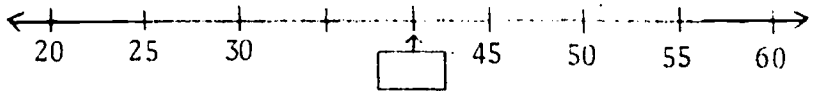
Subtest 5: Computation The crucial directions were, "...figure out what goes in the boxes to make the number sentences true."

Example: ☐ - 49 = 0

Booklet B

Subtest 1: Labelling Number Lines The crucial directions were, "...figure out what number would go in the box on each number line." (Note to reader: no student did both this subtest and Subtest 4 of Booklet A in which the number line was not equal-interval.)

Example:



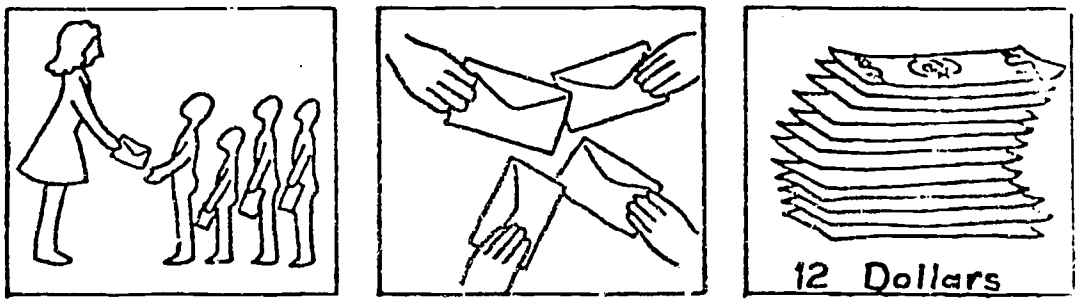
Subtest 2: Number Sentences The crucial directions were, "Write number sentences for 8...write as many as you can think of."

Example: 8 = _____

Subtest 3: Word Problems The crucial directions were, "For each series of pictures there is a story. At the end of each story there is a question you will be expected to answer."

Example: "First picture, 'Four children each get the same allowance from their mother.' Second picture, 'The four children put their allowances together.' Third picture, 'They have altogether 12 dollars.' Question, 'How much did each child get?'"

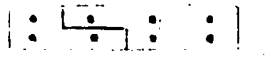
(Note to reader: while the above information was being read to the students, the student's test page contained the pictures below.)



How much did each child get? _____

Subtest 4: Number Sentence Pictures This was a "matching" task; four equations on one side of the page and five dot pictures on the other side. The crucial directions were, "figure out which dot picture shows (each) number sentence best...there are five pictures and only four number sentences...make up a number sentence that goes with the picture that's left over."

Example: 5+3=8



Subtest 5: Computation (Same directions as Subtest 5 in booklet A but different items.)

MANS Results

At each of four participating sites all the CSMP classes and an equivalent number of Non-CSMP classes were tested. The Non-CSMP classes were selected so as to provide a group roughly equivalent (with respect to general academic ability) to the group of CSMP classes. The same two booklets of the MANS were given at each of the sites, with roughly half the CSMP and half the Non-CSMP classes getting booklet A and half of each group booklet B. The testing plan in the various sites is shown in Table 1, page 2.

In order to compare test performance of CSMP vs. Non-CSMP classes an analysis of covariance on the class means was performed across sites. The mean scores at each site, adjusted for the covariate, are shown for CSMP and Non-CSMP classes in Table 4 below. Also shown (right hand columns) are the mean adjusted scores combined across sites and the resulting p-values for each of the MANS tests.

Table 4

Adjusted Mean Scores
By Site and MANS Tests
(First Entry: CSMP Classes, Second Entry: Non-CSMP Classes)

	Site[1] 01	12	13	25	Combined	p-value
MANS A						
Covariate	58.5 56.9	50.4 45.1	62.9 68.2	49.2 42.8	56.1 54.0	
1. Sequences	2.6 2.2	2.4 2.7	2.8 2.3	3.8 2.9	2.9 2.5	.09
2. Equation Fluency	3.5 3.4	2.9 3.2	3.9 3.1	4.7 4.0	3.7 3.4	.43
3. Functions	1.8 1.2	1.8 1.4	2.0 2.2	2.3 1.7	2.0 1.6	.09
4. Number Line Estimation	4.5 3.8	3.3 4.0	4.2 4.2	5.7 4.6	4.4 4.1	.23
5. Computation I	6.4 5.0	7.2 5.7	7.2 5.0	7.4 6.8	7.0 5.6	.01
TOTAL	18.7 15.6	17.6 17.2	20.1 16.8	23.9 19.9	20.0 17.4	.04
MANS B						
Covariate	59.6 54.5	64.7 41.3	61.3 77.2	40.9 41.8	56.7 53.3	
1. Number Line Labelling	4.4 4.0	4.7 6.3	4.7 4.1	6.2 6.2	5.0 5.1	.88
2. Sentences About 8	3.7 2.9	3.1 3.9	3.4 1.9	4.8 5.0	3.7 3.4	.19
3. Word Problems	4.0 3.8	4.2 4.0	4.5 3.7	4.7 4.5	4.3 4.0	.08
4. Number Sentence Pictures	4.8 4.2	5.3 4.7	5.6 3.7	6.5 5.6	5.6 4.5	.01
5. Computation II	6.0 4.8	6.7 5.9	6.7 3.5	8.1 7.3	6.9 5.4	.01
TOTAL	22.9 19.7	23.9 25.0	24.9 16.8	30.3 28.6	25.5 22.5	.02

[1]Number of classes taking MANS A in sites 01, 12, 13 and 25 was 15, 6, 6 and 6 respectively; at each site approximately half were CSMP classes and half were Non-CSMP classes.
For MANS B, the number of classes was 13, 6, 6 and 6 respectively and they were different classes than those who took MANS A.

A review of Table 4 reveals that there are statistically significant differences in favor of CSMP classes ($p<.05$) for the total score on MANS A and on MANS B and for three of the individual subtests: Computation I, Number Sentence Pictures and Computation II. On three other subtests — Sequences, Functions and Word Problems — the differences were nearly significant ($p<.10$). When examining the scores for the combined sites, it can be seen that all differences, whether significant or not, are in favor of the CSMP classes. This pattern is replicated at most sites, although four of the twelve scores at site 12 are in favor of Non-CSMP classes (though not significant).

Test performance as a function of ability level was examined for each subtest by grouping the students into quintiles according to student ability as defined by covariate score. For example, a mean score was computed based on all CSMP students whose covariate score corresponded to a percentile rank less than 20 (lowest quintile). Similar mean scores were computed based on CSMP students with covariate percentile rank from 20 to 40 (second quintile), 40 to 60 (third quintile), 60 to 80 (fourth quintile) and 80 or more (highest quintile). Then the same thing was done for Non-CSMP students. The graphs in Figure 1, below, allow visual comparisons of CSMP and Non-CSMP students at each quintile, and one can judge the degree to which overall higher scores by CSMP students is caused by students at a particular ability level.

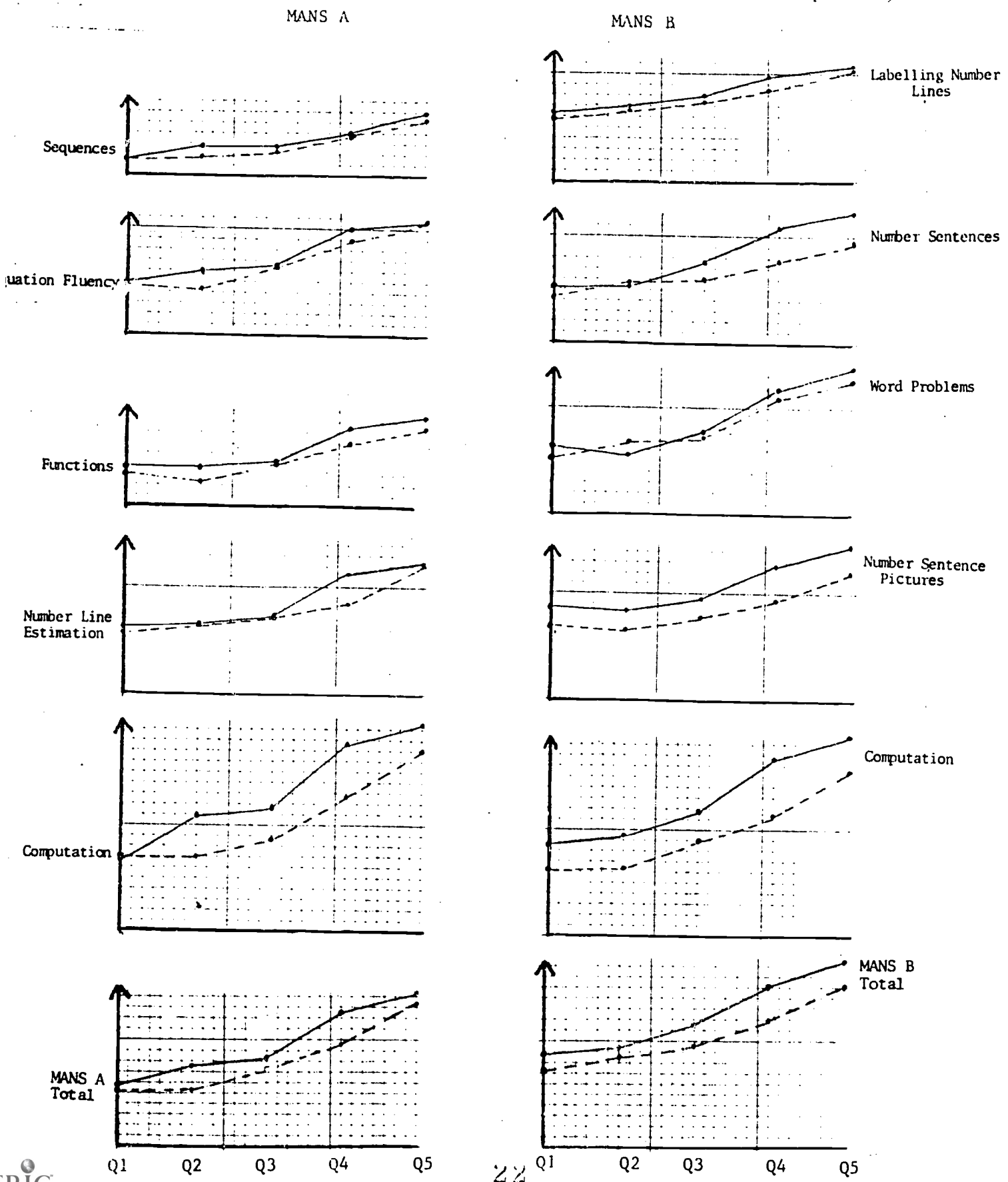
It can be seen in Figure 1 that on each subtest the CSMP mean for the various quintiles is almost always larger (higher on the graph) than the corresponding Non-CSMP mean. In other words CSMP students are doing better than Non-CSMP students; not just overall, but at each ability level. It is at the fourth quintile (percentile rank 60-80) that the largest and most consistent differences occur.

Appendix G also presents an item analysis of the MANS tests, with percent correct for each item across CSMP and Non-CSMP students. Included with the item analysis are graphs showing the distribution of class means for each MANS test according to class ability (covariate) scores, and various other summary statistics.

Figure 1

CSMP (—) and Non-CSMP (- -) Means by Quintile

(The vertical axis is for raw score, with one unit equal to one half of a raw score point, except for the two total scores where one unit equals two raw score points.)



Discussion

After a brief treatment of the data obtained from the CSMP-Specific tests, the bulk of this discussion will focus on the data resulting from CSMP - Non-CSMP comparisons.

Although a great deal of information is available from the various CSMP-Specific Tests, the spiral nature of the curriculum and consequent lack of mastery levels at any given point in time make it difficult to assess the adequacy of student performance in this area. It does appear for the most part that students were moderately successful in doing the workbook-level pages covered in these tests. (On the second grade CSMP-Specific Tests A and B for example, with the exception of the subtest "Integers", the proportion of correct answers was between 2/3 and 3/4 for each subtest. In fact, of the total of 60 test items, the percent correct was between 70 and 90 for 62% of them.) However, it is clear that proficiency with Minicomputers and arrow diagrams are not particularly important objectives per se. What is important to know is how well a curriculum with these pedagogical devices succeeds in having children learn mathematical information (such as numerical concepts and computational skills) and learn to apply this information usefully.

With regard to comparisons between CSMP and Non-CSMP students the pattern of results is very similar to the pattern in previous years of the Extended Pilot Test, with one very important exception: these results are from a much broader base of classes in terms of geographic location. In particular, these results have been extended to sites outside the St. Louis area and are thus free from the certain non-generalizable features such as teacher training by CSMP personnel, knowledge of being a "horserace", frequent observations from visitors, etc.

Briefly one can summarize the results as follows:

a) On standardized mathematics tests, CSMP classes generally did as well as or better than Non-CSMP classes. At third grade the differences were significant on both tests of the CTBS. At second grade the results were more equivocal; not reaching significance at any of the four sites, but usually in favor of CSMP, especially on the tests of computational skills. Similar results have occurred in previous years; small differences in favor of CSMP students which sometimes reached significance and sometimes did not.

b) On the MANS tests, CSMP classes scored significantly higher than Non-CSMP classes on both total MANS scores and on three of the subtests, and were close to significantly higher ($p<.10$) on three other subtests. Again this pattern is rather similar to results from previous years: consistently higher scores by CSMP classes in all subtests with significance reached on certain subtests and on the overall totals.

The value one attributes to the superior performance on the MANS tests is of course dependent on the importance one gives to the mathematical abilities and thinking processes required by them. It is true that these tests are biased to the extent that they were developed with the CSMP curriculum and pedagogy in mind and that a different set of tests might have been developed with entirely different results. Nevertheless the actual tasks of the MANS tests were usually quite different from those typically found in either the CSMP curriculum or in most elementary school math texts; they were novel to both groups and used

none of the special languages of CSMP. Such transfer effects are relatively unusual in curriculum evaluation.

It is also true that certain mathematical content embedded in these tasks may favor the CSMP classes. If one looks at the percentage correct for CSMP and Non-CSMP classes on an item-by-item basis, one finds that several of the items on which there are substantial differences in favor of CSMP have to do with: a) multiplying, b) using fractions or c) working with "larger" numbers (in the hundreds). CSMP does stress these ideas somewhat more than most math programs. Thus, at least some of the differences between CSMP and Non-CSMP classes on the MANS test may be due to a difference in the mathematical content covered by the respective curricula. However, whether one views it as knowing more content or as being better able to use certain processes, the gains for CSMP seem to have been achieved without any compensating loss in other areas (at at least none that showed up on the standardized measures).

Given for the moment that the favorable outcomes of the CSMP - Non-CSMP comparisons are accepted, it is necessary to consider alternative explanations for these outcomes before concluding that the CSMP curriculum made the difference. Probably the most plausible of these is that CSMP teachers as a whole may have been a better (or more enthusiastic) group of teachers than were Non-CSMP teachers. They may have been selected (or volunteered) because of certain favorable attributes, or may have developed same as a result of this selection (or volunteering). There is no way to adequately test this possibility after the fact; research has shown that objective criteria of teacher attributes generally do not bear a consistently strong relationship to student achievement. The use of a design whereby volunteer teachers are randomly assigned to CSMP and to Non-CSMP classes was not practical in this study, nor is it generally practical in educational settings.

(Now it is true that the explanation of teacher differences is most plausible during the first year of a study, when teachers may well have volunteered; in subsequent years teachers to some degree "inherit" the CSMP class from the previous year - though some selection still may take place. It is also true that a previous study found that teachers in their second year of CSMP did at least as good a job of teaching CSMP as they did in their first year, as measured by student achievement (see Evaluation Report 2-B-2). Thus any decrease in teacher enthusiasm, at least from first to second grade, did not appear to affect student achievement. These considerations tend to reduce, but not eliminate, the hypothesis that teacher-differences are responsible for CSMP - Non-CSMP differences.)

Among CSMP classes there is considerable variation in: a) the amount and kind of training received by teachers, b) the amount of time per day devoted to mathematics instruction and c) the amount of supplementary materials, such as drill sheets, which are assigned to students. It is not clear how much corresponding variation in achievement, if any, results from each of these factors. From an experimental point of view, they may be regarded as rather natural variations in the implementation of CSMP in the "real world". And there are similar variations in Non-CSMP classes. (For example, most teachers, no matter what curriculum they are using, routinely assign supplementary material of some kind; some more than others.) It is for this reason that one uses classes as the unit of analysis; a more conservative but realistic method than using individual students.

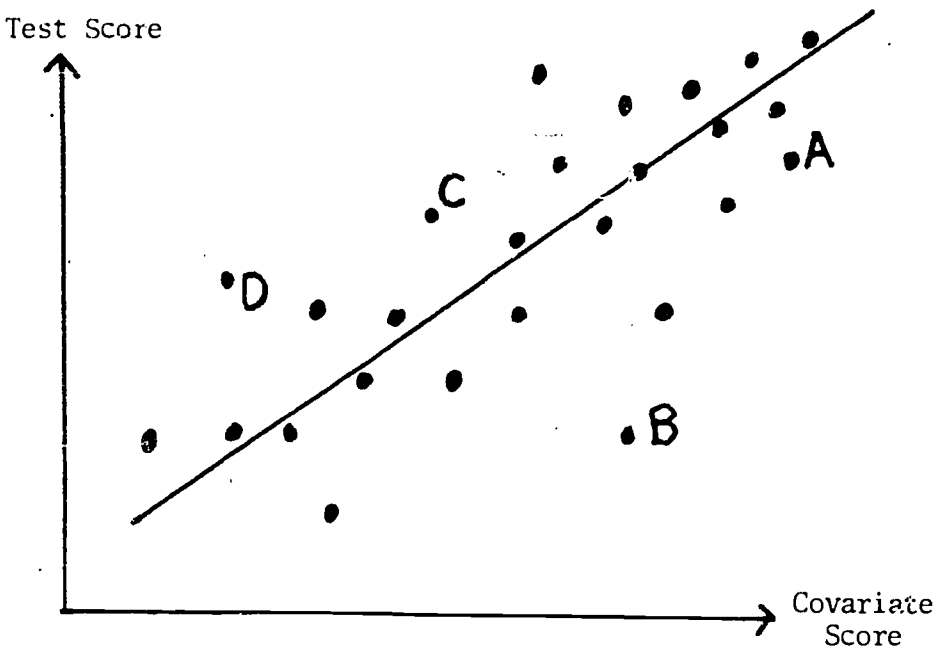
Finally one should note that considerable credibility is added to the results by their consistency. The results have been fairly consistent from year-to-year, across different levels of student and class ability, and in different educational settings and school districts. Bearing in mind the potential effects of teacher selection, these results offer considerable promise for the CSMP curriculum.

Introduction to Appendices

Two methods of data presentation have been used in various appendices. One method involves aggregating across students to get percentages correct on individual test items; the other involves the calculation of mean scores for each class and then using the set of class mean scores to carry out tests of significance. These are described in more detail below.

First, the actual test items making up a particular test are given. Beside these items are given percentage correct for all CSMP students and for all Non-CSMP students (where the test was also given to Non-CSMP students). Occasionally, where covariate scores (reading or general ability scores) were available, this analysis was also carried out by level of student ability. This was done by transforming covariate scores to percentile rank and doing separate analyses for the set of students with percentile rank less than 20 (lowest quintile), the set of students with percentile rank 20-39 (second quintile), and so on to the set of students with percentile rank of 80 or more (fifth or highest quintile).

Second, when class means are being analyzed for some particular test, a mean score is calculated for each class on that test and on the covariate test. Then, by transforming the covariates to percentile ranks, each class can be represented by a dot, as illustrated below.



The vertical axis is for mean class test score on whatever test is being analyzed; the further towards the top of the graph - the higher the class score on that test. The horizontal axis is for covariate score, or mean class score on whatever covariate was used; the further to the right - the higher the general ability of the class. Based on the set of class means thus graphed, the regression line has been drawn. This line is the best linear prediction of mean class test score that can be made from knowing the ability level of the class. Note in the example that classes A and B fall well below the regression line, or are scoring well below what would be predicted for them from knowing the ability level of the class, while classes C and D fall well above the regression line. Note also that, although class A had a slightly

higher mean score on the test than did class D, class D did much better given the relative ability scores of the two classes. When the class means generally fall close to the regression line, test scores are well predicted by the covariate; when they are more dispersed from this line, the covariate is a less effective predictor.

For tests in which CSMP - Non-CSMP comparisons were made, summary statistical data has been shown in some unused corner of the graph. The method of comparison was the Analysis of Covariance, using class mean as the unit of analysis. Essentially this method assumes that at least part of the test score differences among classes is due to differences in general ability levels of the class (as measured by the covariate) and test scores are adjusted to take into account such differences. The p-value shown at the bottom of these tables is essentially the probability that, after taking into account differences in class ability levels (covariate), any observed differences between CSMP and Non-CSMP class scores are random "errors" from the true state of affairs - which is "no differences between CSMP and Non-CSMP." A low p-value indicates that this is an unlikely explanation of the true state of affairs; that in fact there probably are real differences between CSMP and Non-CSMP classes. Traditionally a p-value less than .05 is interpreted as "significant".

It should also be noted that all analyses, whether aggregated across students or across classes, are based on students who were also present for the administration of the covariate test, where such was given (i.e. where comparisons were made). This usually elininated very few students from the analysis and tended to do so in about equal proportions from CSMP and from Non-CSMP classes.

For the CSMP-Specific Tests (Appendices D, E and F) considerable use of color was made on the actual tests. The reproductions given here are in black and white, with various "thicknesses" of lines taking the place of different color. The results are fairly easily understood, though not an aesthetic success.

At the end of most Appendices, a brief Commentary is given, in which various points of interest (at least to the authors) are noted. To aid the reader, subsequent pages are labelled with both the page number and the letter for the Appendix (i.e. "115" means page 115 of the report and part of Appendix F).

Appendix A

Comprehensive Test of Basic Skills, Second Grade, Site 01

There are two mathematics tests in the Comprehensive Tests of Basic Skills: Computation, and Concepts and Applications. For the purpose of this analysis, the Computation Test has been subdivided into three sections: addition items, subtraction items and multiplication items. These sets of items are also separate sections in the test but are ordinarily combined to form a gross Computation score.

For each subtest, two kinds of information are given. First, a reduced version of the test items is given (usually on the left hand page) together with percent correct for CSMP students and for Non-CSMP students. An asterisk indicates that the difference in percent correct was greater than 10. All such differences favored CSMP students.

Second (usually on the right hand page), a graph of class means is given which corresponds to the set of test items on the facing page. CSMP classes are shown by an x and Non-CSMP classes by a dot (●). Also shown on each graph is a statistical summary of the class mean data, including the test of significance for differences between CSMP and Non-CSMP classes.

Included in the analysis were 272 CSMP students with a mean raw score on the Kuhlmann-Anderson Test (administered the previous fall) of 48.2 and 251 Non-CSMP students with a mean score of 46.3.

The information is presented in the following order:

addition items
subtraction items
multiplication items
Computation test (sum of the above sets of items)
Concepts and Applications Test

CTBS Item Analysis, Second Grade, Site 01

Subtest 1: Addition (10 items)

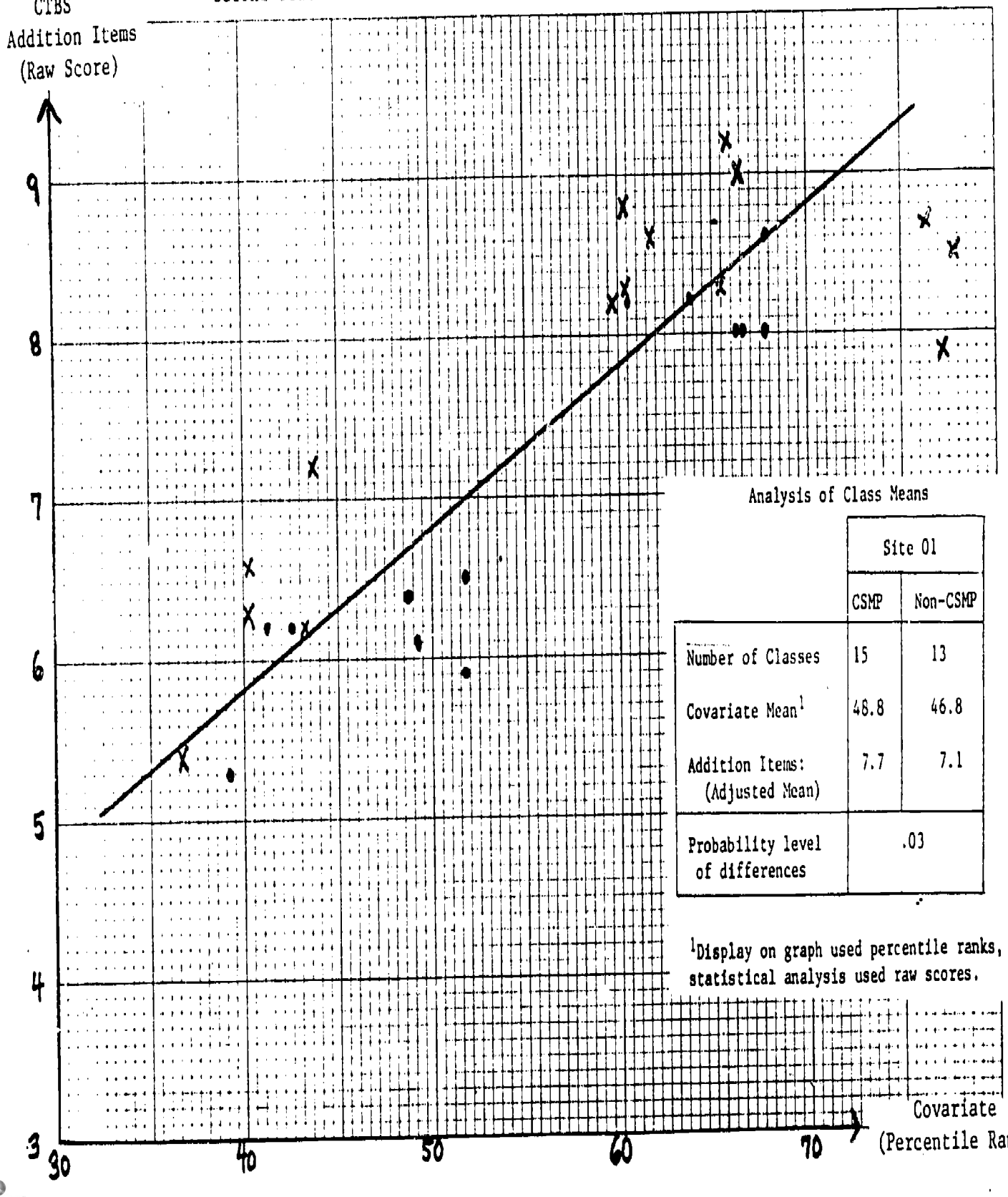
	Percent Correct			Percent Correct	
	CSMP	Non-CSMP		CSMP	Non-CSMP
3 8 0 + 6	87	89	13 6 52 + 27	64	49 *
19 + 6	83	76	40 + 30 =	91	87
154 +128	80	64 *	7 + 3 + 2 =	88	91
27 +18	68	57 *	11 + 7 =	90	88
1.14 + .63	41	28 *	26 + 21 =	83	69 *

23 A

Site 01

CTBS (addition items) versus Covariate (Kuhlmann Anderson)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

	Site 01	
	CSMP	Non-CSMP
Number of Classes	15	13
Covariate Mean[1]	48.8	46.8
Addition Items: (Adjusted Mean)	7.7	7.1
Probability level of differences	.03	

[1]Display on graph used percentile ranks, statistical analysis used raw scores.

CTBS Item Analysis, Second Grade, Site 01

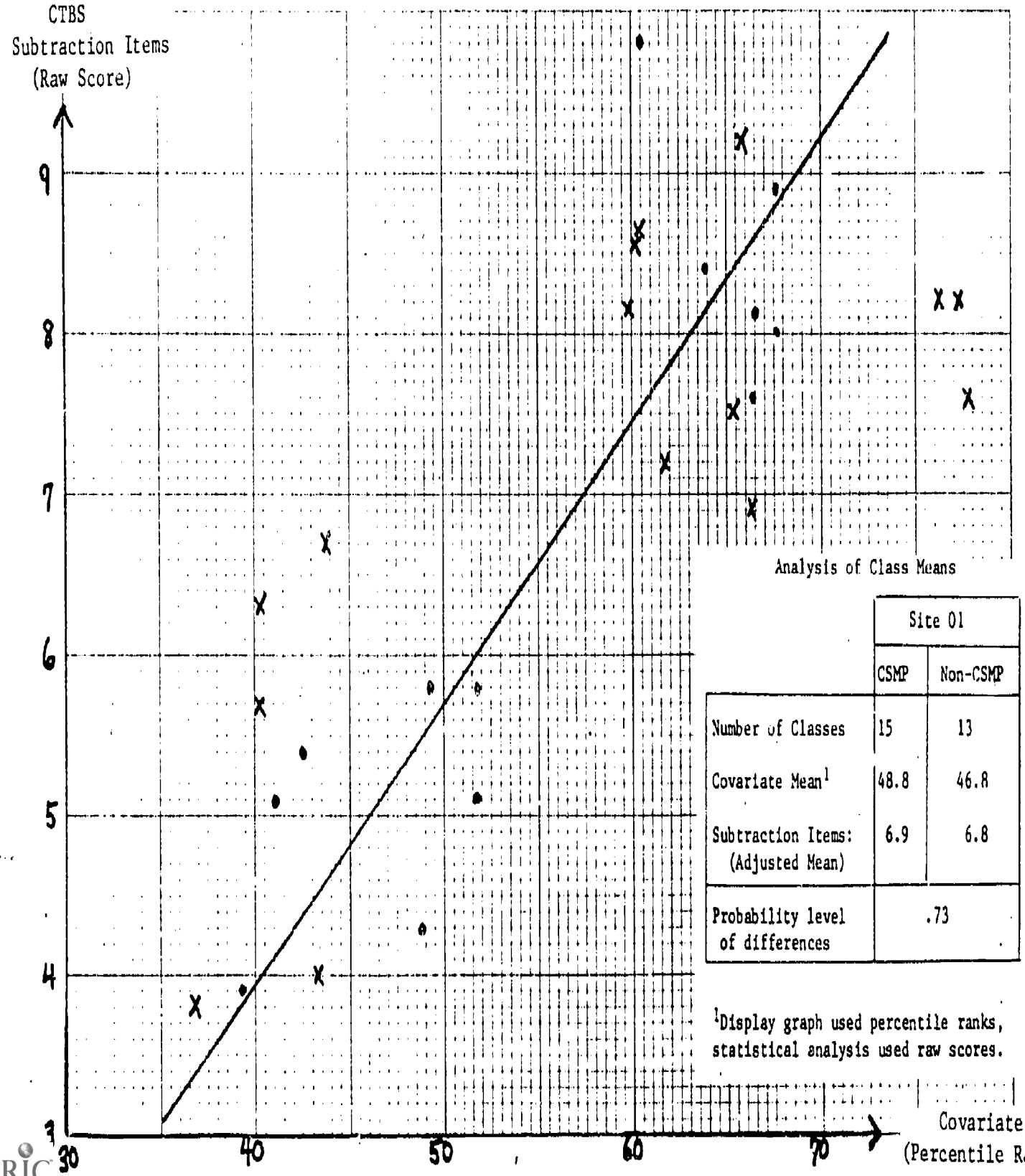
Subtest 2: Subtraction (10 items)

	Percent Correct			Percent Correct		
	CSMP	Non-CSMP		CSMP	Non-CSMP	
38 − 6	79	78	64 − 9	68	57	*
545 − 33	71	64	65 − 42 =	80	75	
16 − 8	72	73	24 − 8 =	71	67	
768 −427	74	60 *	13 − 8 =	66	69	
55 −22	65	61	78 − 43 =	64	54	

Site 01

CTBS (subtraction items) versus Covariate (Kuhlmann Anderson)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

	Site 01	
	CSMP	Non-CSMP
Number of Classes	15	13
Covariate Mean[1]	48.8	46.8
Subtraction Items: (Adjusted Mean)	6.9	6.8
Probability level of differences	.73	

[1]Display graph used percentile ranks, statistical analysis used raw scores.

CTBS Item Analysis, Second Grade, Site 01

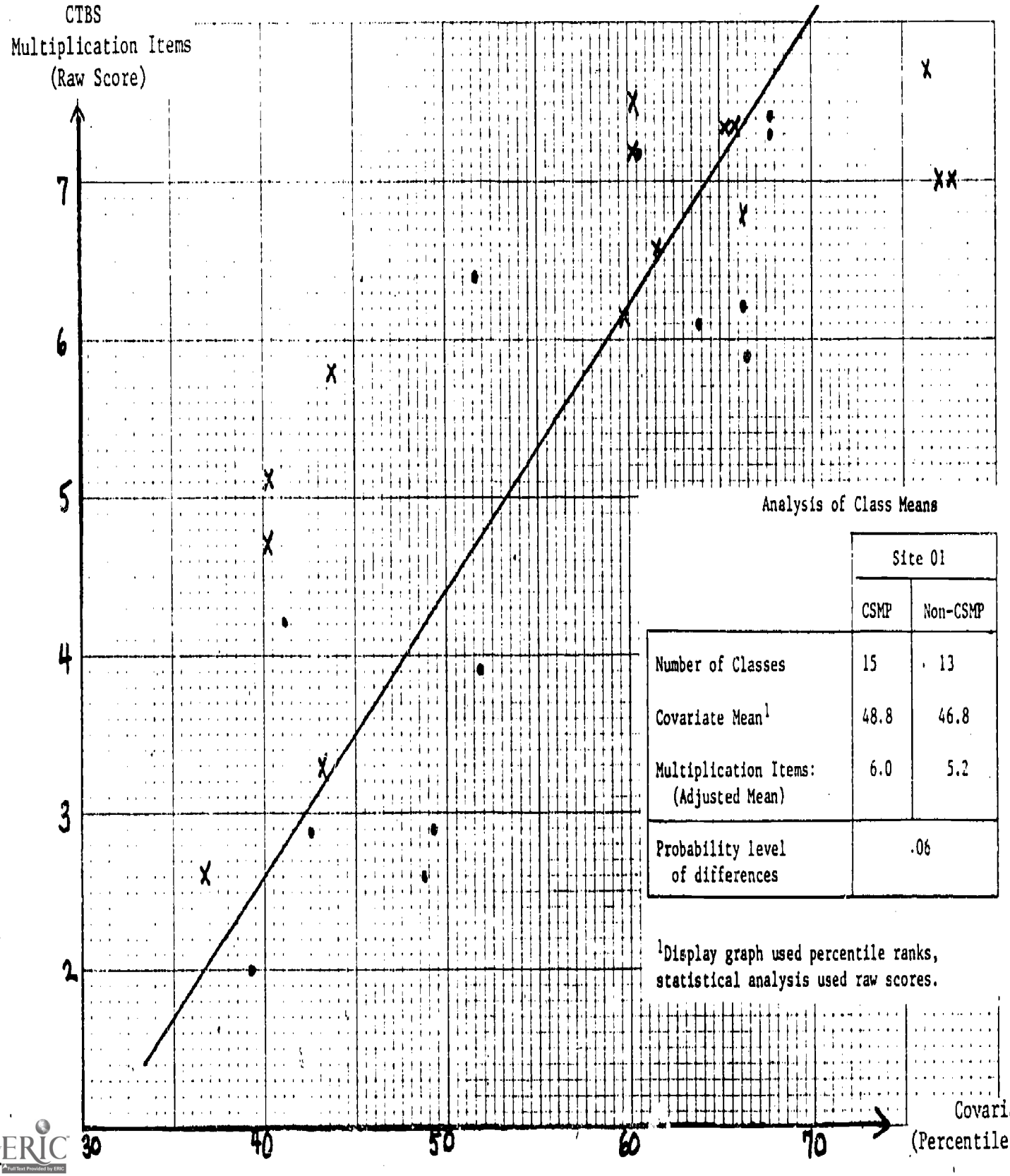
Subtest 3: Multiplication (8 items)

	Percent Correct				Percent Correct		
	CSMP	Non-CSMP			CSMP	Non-CSMP	
5 x 2 =	82	68	*	3 x 4 =	81	64	*
1 x 4 =	83	67	*	5 x 9 =	64	58	
2 x 3 =	83	64	*	4 x 4 =	71	52	*
3 x 5 =	81	65	*	8 x 5 =	71	55	*

Site 01

CTBS (multiplication items) versus Covariate (Kuhlmann Anderson)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

	Site 01	
	CSMP	Non-CSMP
Number of Classes	15	13
Covariate Mean[1]	48.8	46.8
Multiplication Items: (Adjusted Mean)	6.0	5.2
Probability level of differences	.06	

[1]Display graph used percentile ranks, statistical analysis used raw scores.

CTBS, Second Grade, Site 01

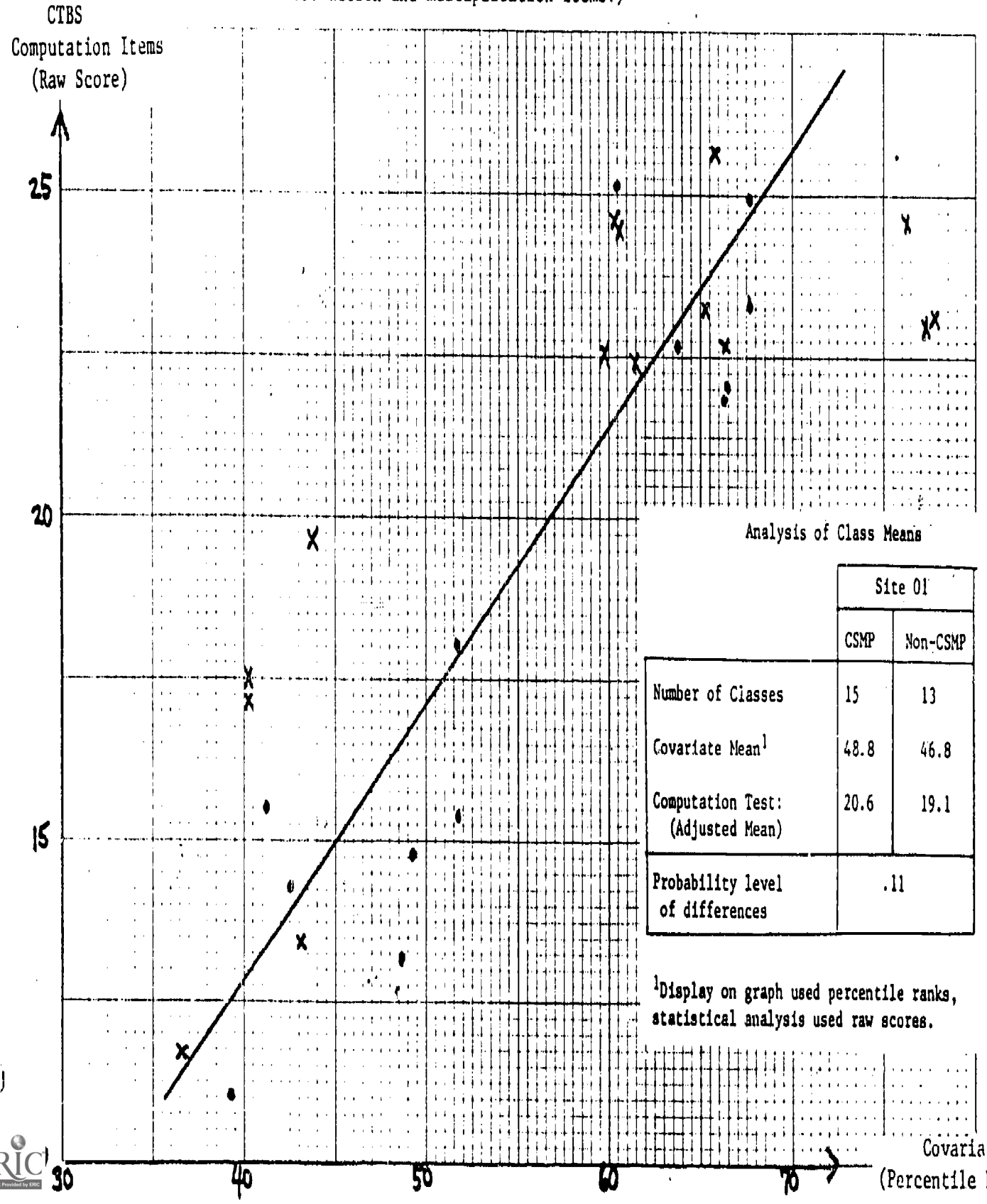
The graph on the next page is for the Computation Test which is composed of the three previous sets of items for addition, subtraction and multiplication.

Site 01

CTBS Computation versus Covariate (Kuhlmann Anderson)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)

(Note: Computation Test is composed of the addition, subtraction and multiplication items.)



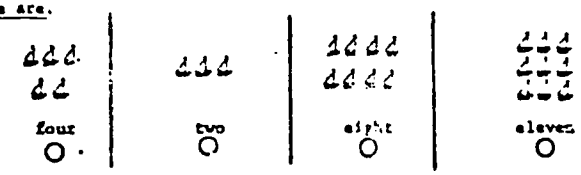
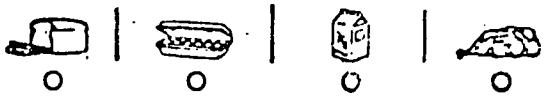
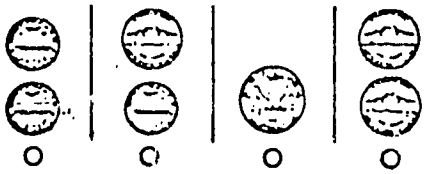
Analysis of Class Means

	Site 01	
	CSMP	Non-CSMP
Number of Classes	15	13
Covariate Mean[1]	48.8	46.8
Computation Test: (Adjusted Mean)	20.6	19.1
Probability level of differences	.11	

[1]Display on graph used percentile ranks, statistical analysis used raw scores.

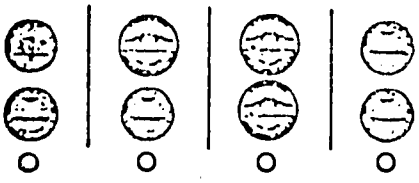
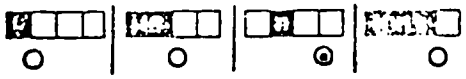
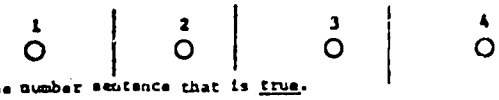
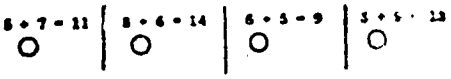
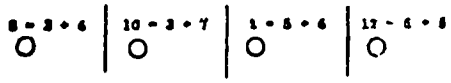
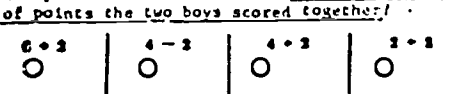
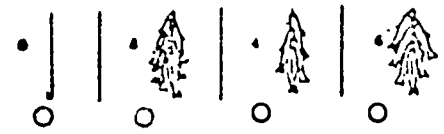
CTBS Item Analysis, Second Grade, Site 01

Subtest 4: Concepts and Applications (25 items)

	Percent Correct	
	CSMP	Non-CSMP
1. Find a box that has a way of showing thirteen. (11 + 2 | 12 + 2 | 11 + 1 | 10 + 2)	84	87
2. Find the box where the number word tells how many boats there are. (four | two | eight | eleven) 	78	75
3. Find the numeral that shows two hundreds, three tens, and six ones. (236 | 263 | 326 | 362)	82	82
4. Find the box that shows counting by fives, starting at 25. (25, 30, 35, 40 | 25, 28, 31, 34 | 25, 27, 29, 31 | 25, 35, 45, 55)	78	80
5. Find the box that shows counting by twos, starting at 77. Mark (77, 73, 69, 65 | 77, 81, 85, 89 | 77, 78, 79, 80 | 77, 79, 81, 83)	70	60
6. Find the box that shows food that is sold by the quart. 	76	75
7. Mark the circle under the amount of money shown in the picture.  (45¢ | 36¢ | 39¢ | 32¢)	79	73
8. Find the box that shows the same amount of money as a dime. 	89	87

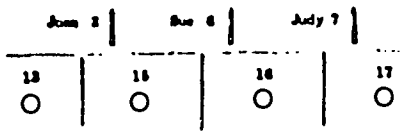
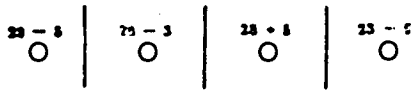
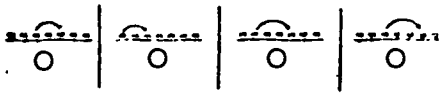
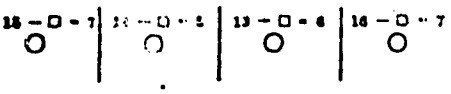
CTBS Item Analysis, Second Grade, Site 01

Subtest 4: Concepts and Applications (cont.)

	Percent Correct	
	CSMP	Non-CSMP
9. David bought a kite that cost nineteen cents. He gave the store clerk a quarter. Find the box that shows how much change David got back. 	59	51
10. Monday is the second day of the week. Find the box that shows what day of the week Thursday is. Mark your answer. 	40	45
11. Find the clock that shows fifteen minutes after seven o'clock. 	72	73
12. Find the clock that shows ten thirty-five. 	64	51 *
13. Find the numeral that is one-half of six. 	82	67 *
14. Find the box that shows that three-fourths of the rectangle is dark. 	90	96
15. Find the numeral that shows one-half of four. 	31	67 *
16. Find the number sentence that is true. 	86	86
17. Find the number sentence that is true. 	82	77
18. Mike scored four points for his team. His brother scored two points. Which box shows how to find the number of points the two boys scored together? 	78	76
19. One day Frank caught six fish. The next day he caught no fish. Find the box that shows how many fish Frank caught all together in the two days. 	84	77
20. Joe invited six boys and eight girls to a party. Find the box that shows how many children he invited all together. 	89	88

CTBS Item Analysis, Second Grade, Site 01

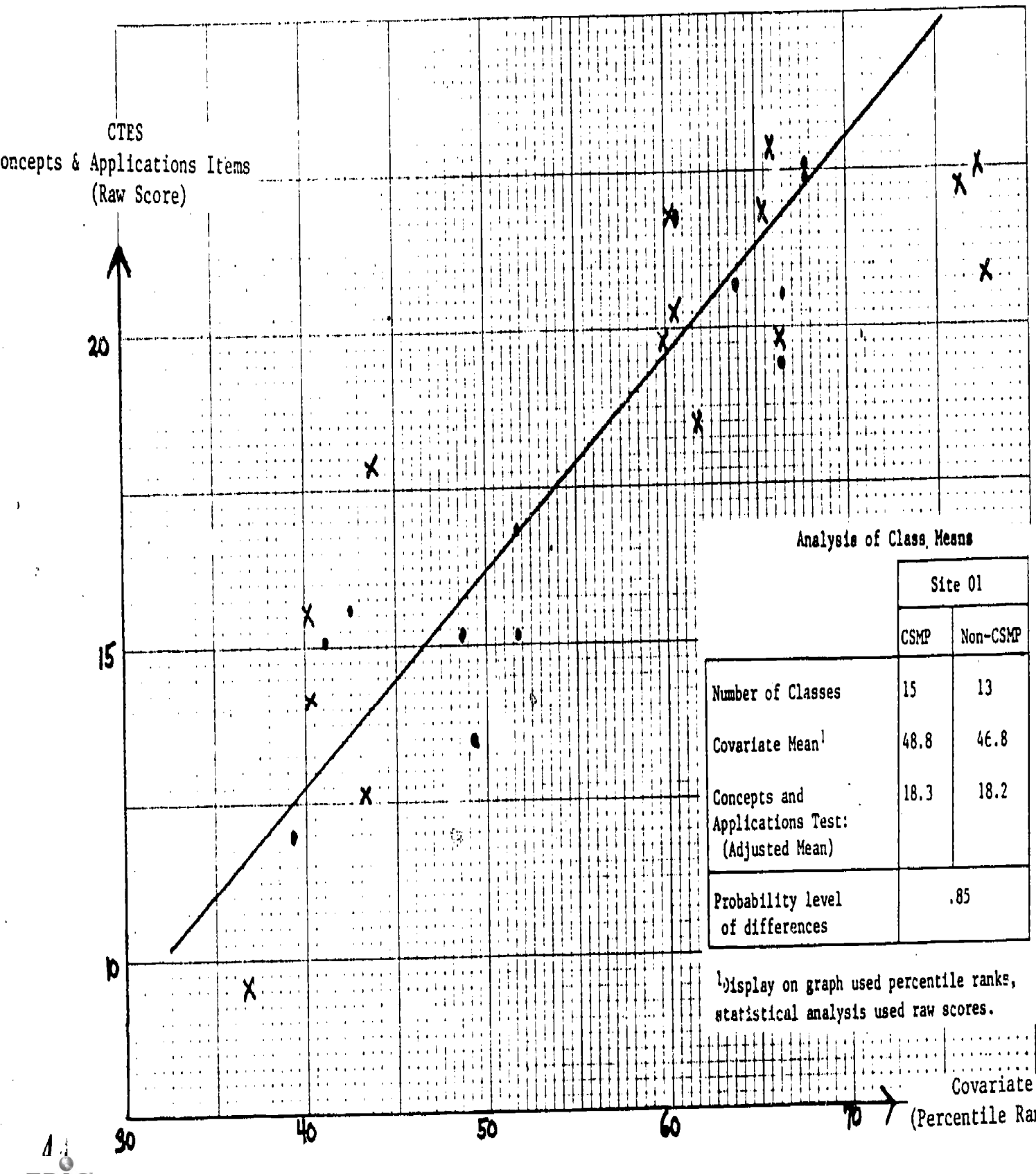
Subtest 4: Concepts and Applications (cont.)

	Percent Correct	
	CSMP	Non-CSMP
21. Joan has three crayons. Sue has six crayons, and Judy has seven crayons. Find the box that shows how many crayons the girls have all together. 	81	84
22. There are twenty-eight children in a class. On Friday, five children were absent. Which box shows how to find the number of children that were at school on Friday? 	61	57
23. Find the number line that shows forty-five plus three. 	71	54
24. Leroy gets up at seven o'clock. He should be at school two hours later. Find the box that shows what time Leroy should be at school. 	61	52
25. Find the box where eight is the missing number. 	53	53

Site 01

CTBS Concepts versus Covariate (Kuhlmann Anderson)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

	Site 01	
	CSMP	Non-CSMP
Number of Classes	15	13
Covariate Mean[1]	48.8	46.8
Concepts and Applications Test: (Adjusted Mean)	18.3	18.2
Probability level of differences	.85	

[1]Display on graph used percentile ranks, statistical analysis used raw scores.

Commentary

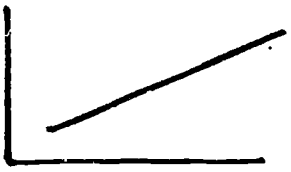
Overall, these results were similar to those obtained in a similar study last year (see Evaluation Report 2-B-1). CSMP and Non-CSMP students got approximately the same score on the Computation Test this year as did CSMP and Non-CSMP second graders respectively last year (after differences in ability were taken into account). The differential in favor of CSMP, about a point and a half each year, was significant last year; not quite significant this year. This may be because of the less restrictive (and less powerful) experimental design in the present study, where Non-CSMP classes were not generally selected from the same school as CSMP classes, which eliminated the possibility of a two-way Analysis of Covariance.

Of the three sets of items making up the Computation Test, this year's CSMP students did relatively better in addition and worse in subtraction than similar students last year. This year's Non-CSMP students did relatively better in addition and worse in multiplication than similar students last year.

On the Concepts and Applications Test, the differential between CSMP and Non-CSMP students was virtually nil this year; almost a full point last year. The difference is due to this year's Non-CSMP group outperforming last year's.

At the item level, the pattern of responses was also similar to last year's. The set of asterisked items in this Appendix, which most differentiated the performance of CSMP and Non-CSMP students, is almost identical to the set of items with the largest differential last year. It is hard to see any particular pattern from this set of items, but the results seem to be consistent.

It is worth noting that the regression model did not fit the set of CSMP class means particularly well for this data; in fact poorer than for any others in this report. A glance at the graphs in this Appendix will reveal that the CSMP class means tend to be curvilinear rather than linear, i.e.

this  , rather than this 

Appendix B

Other Standardized Test Data, Second Grade

The previous Appendix, Appendix A, provided item and summary data from the math tests of the Comprehensive Tests of Basic Skills administered in Site 01. Appendix B provides data from the administration of standardized tests in mathematics at three other sites. In each case class means for CSMP and Non-CSMP classes are plotted on a graph against the class mean covariate scores. Summary statistical information, including adjusted means and the probability level of the differences, is presented in one corner of the graph.

The graphs which are given, and the standardized math tests used are the following:

Site 12: Stanford Achievement Test, Computation Test
Stanford Achievement Test, Concepts Test

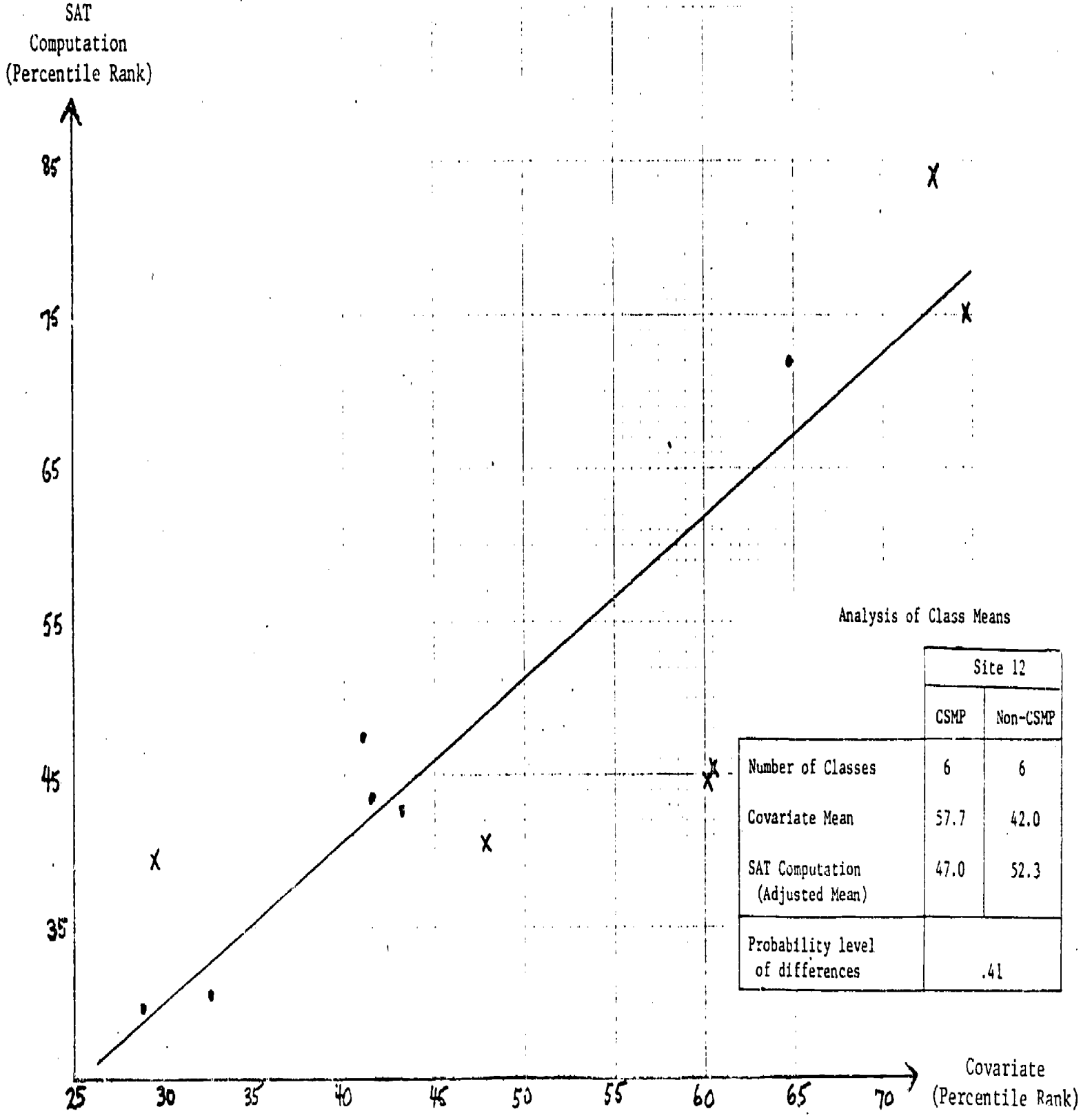
Site 25: Comprehensive Tests of Basic Skills, Computation Test
Comprehensive Tests of Basic Skills, Concepts and Applications Test

Site 13: Cooperative Primary Tests, Mathematics Test

Site 12

SAT Computation versus Covariate (SAT Reading)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



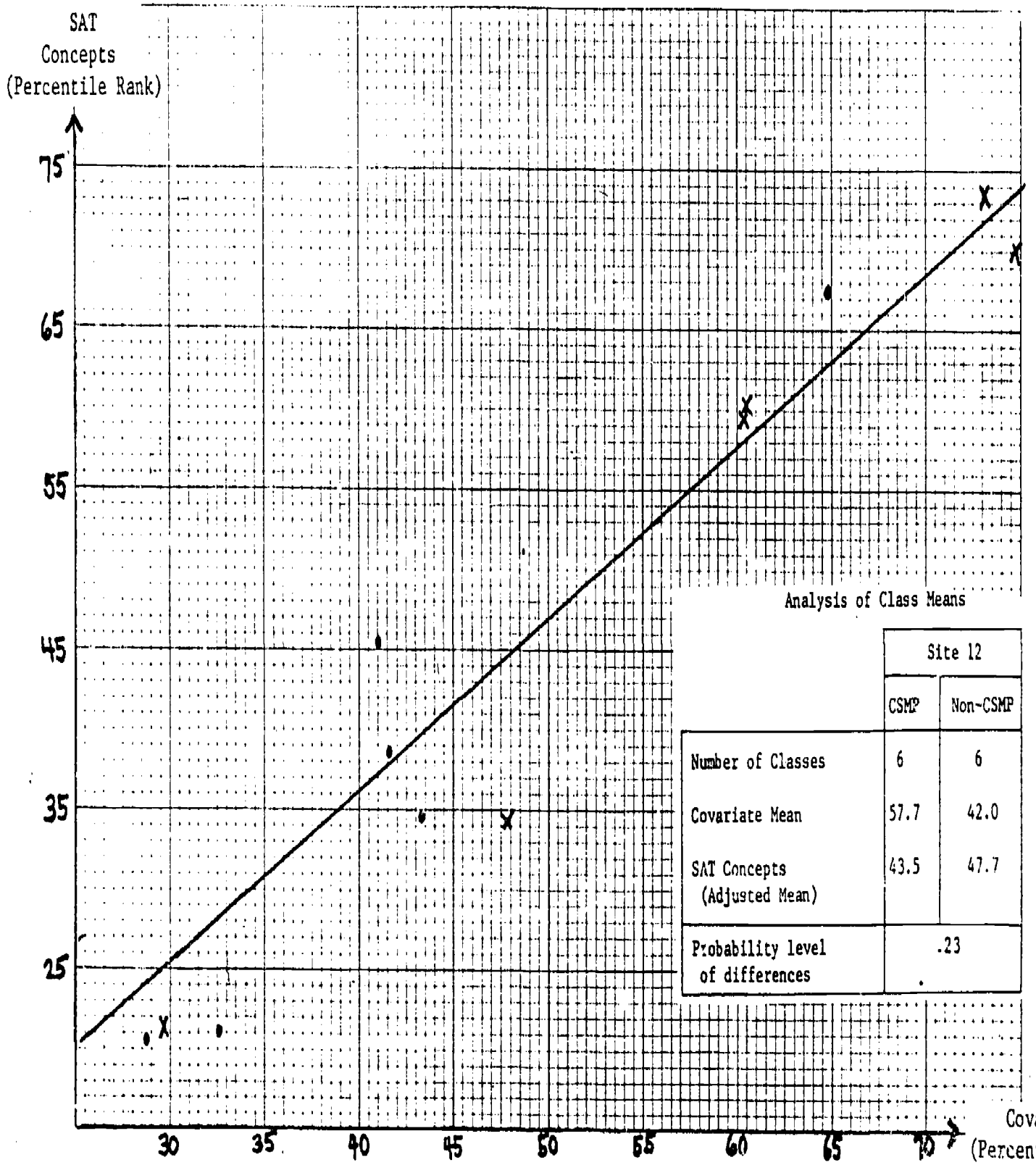
Analysis of Class Means

	Site 12	
	CSMP	Non-CSMP
Number of Classes	6	6
Covariate Mean	57.7	42.0
SAT Computation (Adjusted Mean)	47.0	52.3
Probability level of differences	.41	

Site 12

SAT Concepts versus Covariate (SAT Reading)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



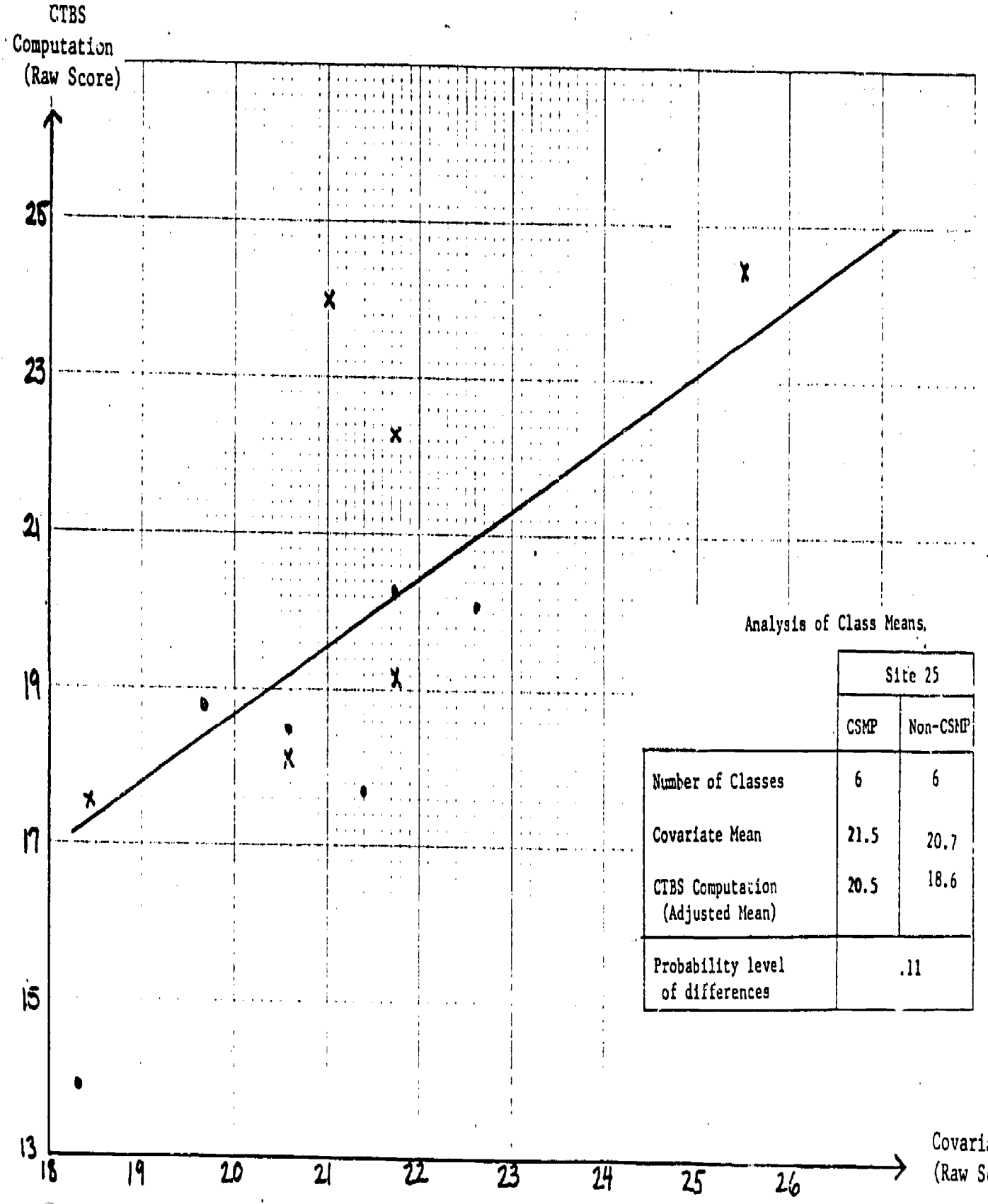
Analysis of Class Means

	Site 12	
	CSMP	Non-CSMP
Number of Classes	6	6
Covariate Mean	57.7	42.0
SAT Concepts (Adjusted Mean)	43.5	47.7
Probability level of differences	.23	

Site 25

CTBS Computation versus Covariate (Gates McGinitie)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



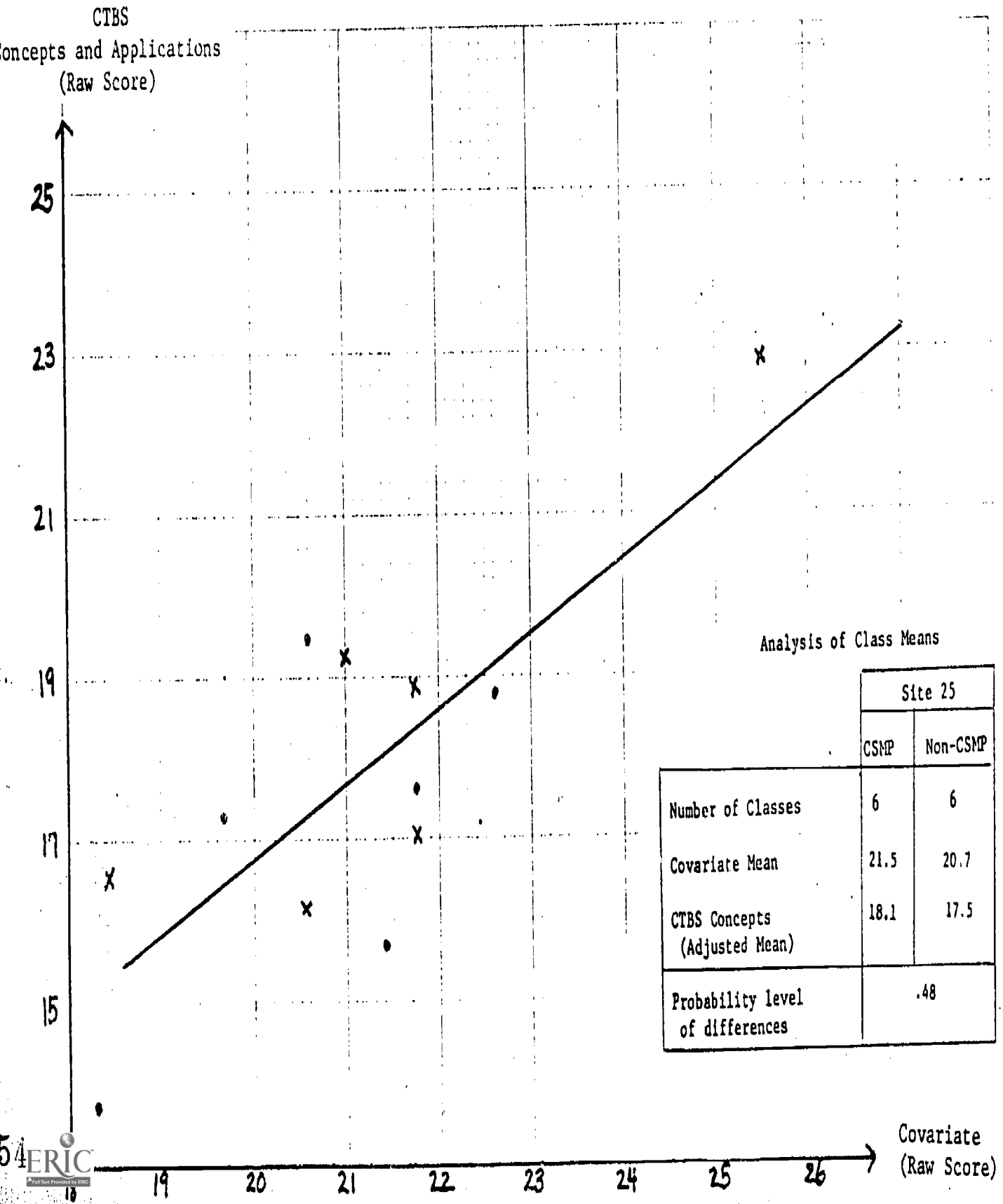
Analysis of Class Means.

	Site 25	
	CSMP	Non-CSMP
Number of Classes	6	6
Covariate Mean	21.5	20.7
CTBS Computation (Adjusted Mean)	20.5	18.6
Probability level of differences	.11	

Site 25

CTBS Concepts and Applications versus Covariate (Gates McGinitie)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



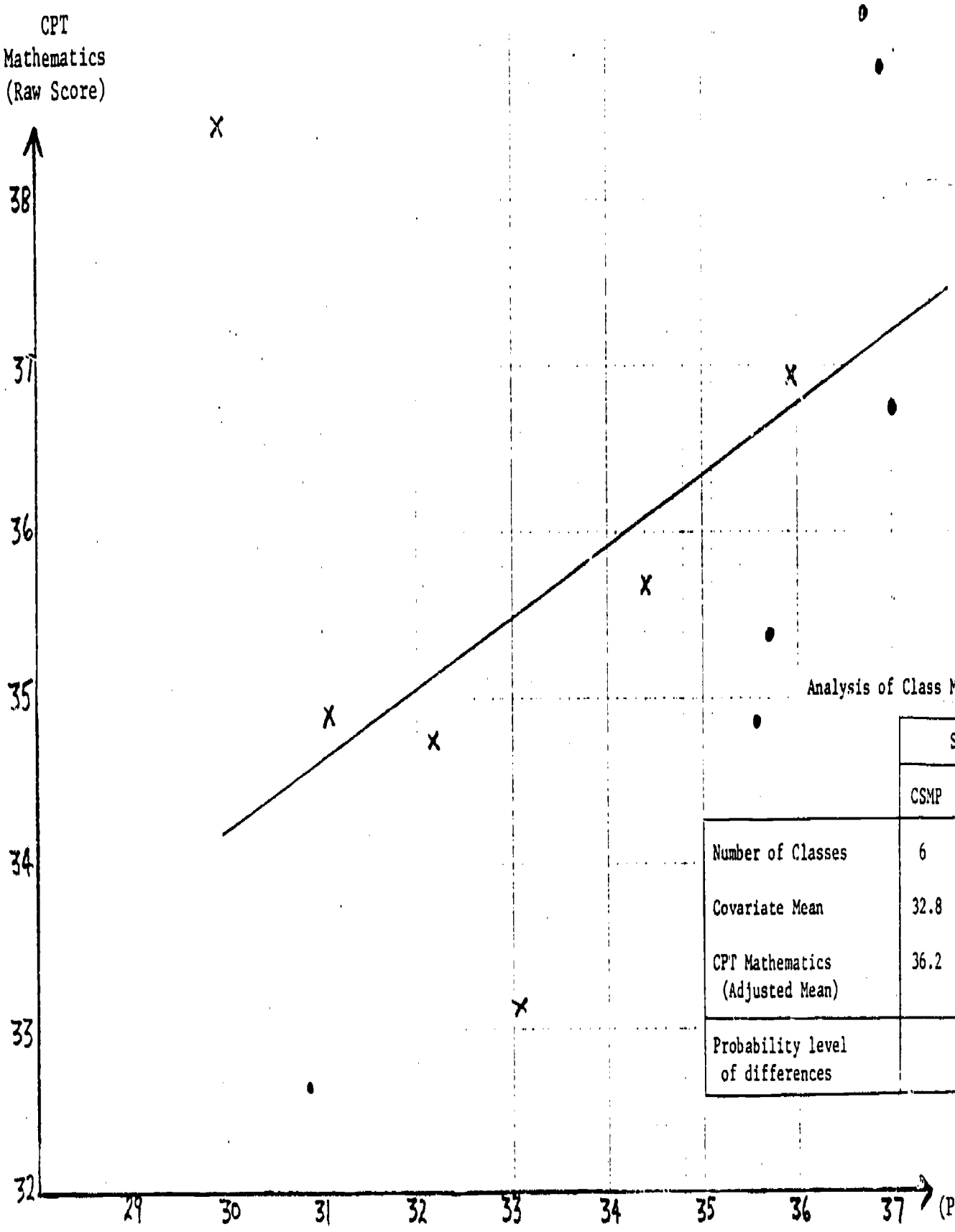
Analysis of Class Means

	Site 25	
	CSMP	Non-CSMP
Number of Classes	6	6
Covariate Mean	21.5	20.7
CTBS Concepts (Adjusted Mean)	18.1	17.5
Probability level of differences	.48	

Site 13

CPT Mathematics versus Covariate (CPT Reading)

Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

	Site 13	
	CSMP	Non-CSMP
Number of Classes	6	6
Covariate Mean	32.8	35.4
CPT Mathematics (Adjusted Mean)	36.2	35.7
Probability level of differences	.18	

Appendix C

Comprehensive Tests of Basic Skills, Third Grade, Site 01

There are two math tests in the Comprehensive Tests of Basic Skills: Computation, and Concepts and Applications. For the purpose of this analysis, the Computation Test has been subdivided into four sections: addition items, subtraction items, multiplication items and division items. These sets of items are also separate sections in the test but are ordinarily combined to form a gross Computation score.

For each subtest, two kinds of information are given. First, a reduced version of each of the various sets of test items are given, together with the percent correct for CSMP and for Non-CSMP students. (This information is on the left hand page.) An asterisk indicates that the difference in percent correct was 10 or more. All such differences favored CSMP students.

Second, a graph of class means is given which corresponds to the previous set of test items. CSMP classes are shown by an x and Non-CSMP classes by a dot (●). (This information is given on the right hand page, facing the corresponding set of items.) Also shown on each graph is a statistical summary of the class mean data, including the test of significance for differences between CSMP and Non-CSMP classes.

There were 148 CSMP students with a mean score of 50.6 on the Kuhlmann-Anderson Test (administered the previous fall) and 148 Non-CSMP students with a mean score of 53.6.

The graphs are presented in the following order:

addition items
subtraction items
multiplication items
division items
Computation Test (Sum of the above sets of items)
Concepts and Applications Test

CTBS Item Analysis, Third Grade, Site 01

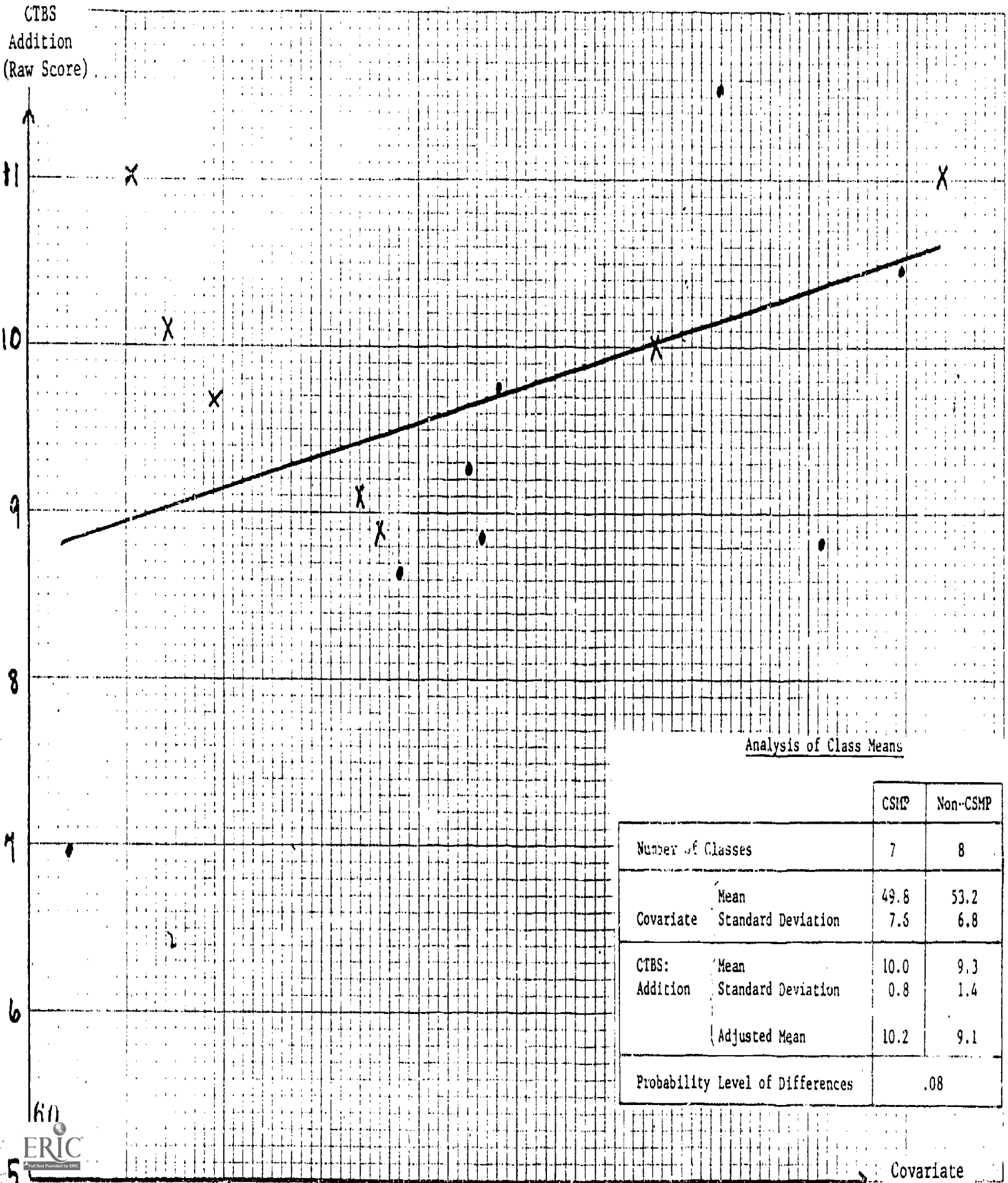
Subtest 1: Addition (12 items)

	Percent Correct			Percent Correct		
	CSMP	Non-CSMP		CSMP	Non-CSMP	
5 6 +4	99	98	730 +593	73	70	
6 0 +7	98	96	32 10 +44	94	84	*
72+73=	96	86 *	26+13+14=	80	65	*
67 +18	85	79	3,536 +7,982	71	62	
79 +14	85	78	35 82 98 + 4	80	74	
$0.56 + 0.68	72	64	44 8,151 432 70 + 108	74	69	

Site 01

CTBS Addition versus Covariate (Kuhlmann-Anderson)

Third Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		7	8
Covariate	Mean	49.8	53.2
	Standard Deviation	7.5	6.8
CTBS: Addition	Mean	10.0	9.3
	Standard Deviation	0.8	1.4
	Adjusted Mean	10.2	9.1
Probability Level of Differences		.08	

CTBS Item Analysis, Third Grade, Site 01

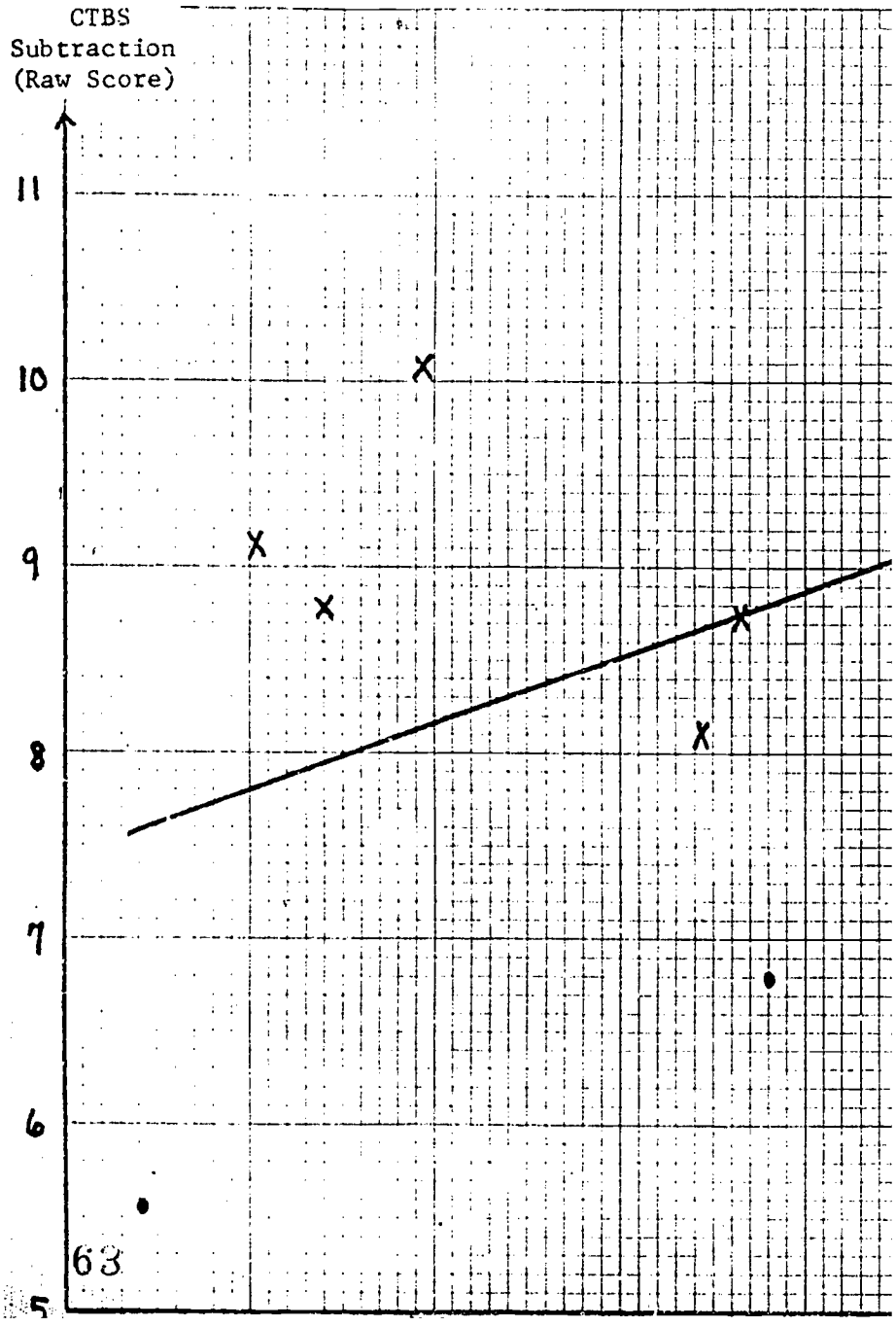
Subtest 2: Subtraction (12 items)

	Percent Correct	
	CSMP	Non-CSMP
97-4=	90	89
33 − 31	87	78
88 − 50	83	84
939 − 18	85	84
136 − 82	75	84
149 − 67	82	72
126 − 92	68	68
300 − 5	62	56
490 − 130	82	87
498 − 203	78	75
564 − 356	68	59
738 − 169	60	51

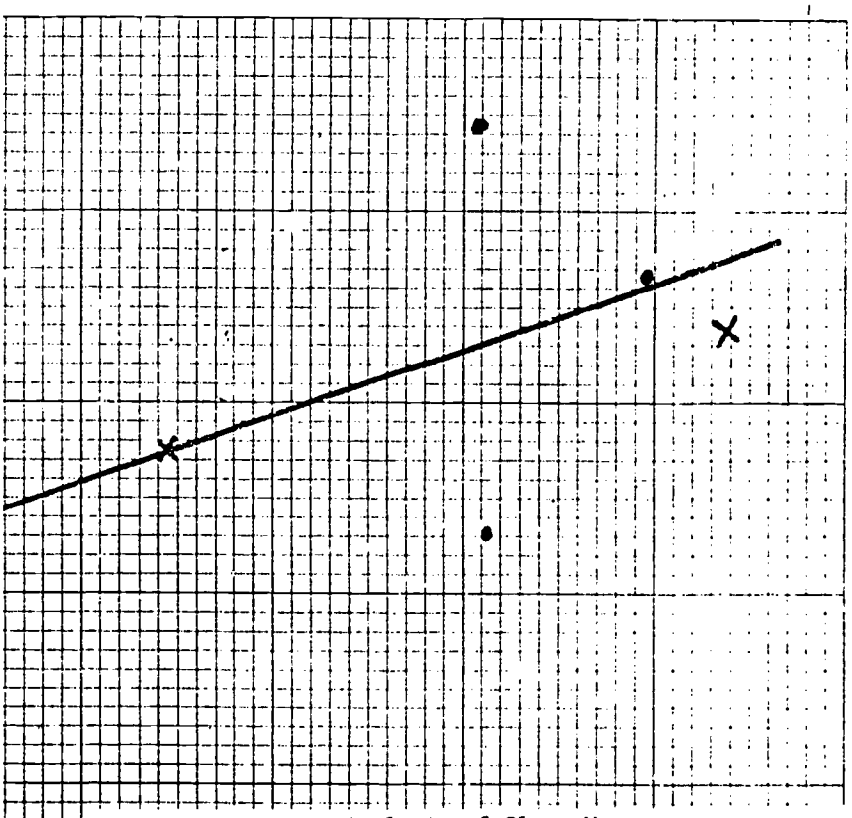
*

45 C

CTBS Subtraction ver
Third Grade Class Means: CS



e (Kuhlmann-Anderson)
) and Non-CSMP Classes (•)



Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		7	8
Covariate	Mean	49.8	53.2
	Standard Deviation	7.6	6.8
CTBS: Subtraction	Mean	9.3	8.9
	Standard Deviation	.8	2.0
	Adjusted Mean	9.5	8.6
Probability Level of Differences		.17	

Covariate

CTBS Item Analysis, Third Grade, Site 01

Subtest 3: Multiplication (12 items)

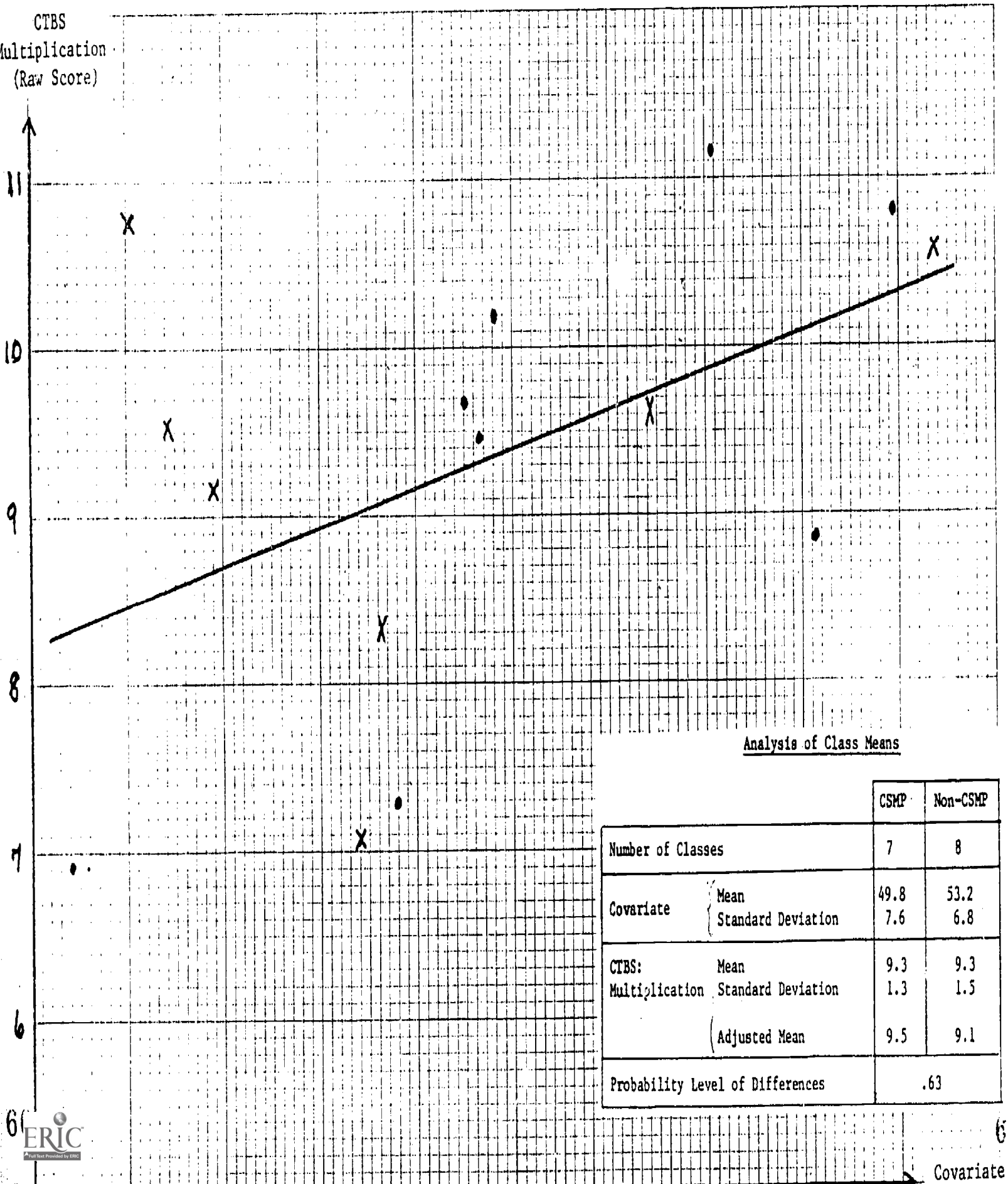
	Percent Correct			Percent Correct	
	CSMP	Non-CSMP		CSMP	Non-CSMP
3x3=	87	96	25 x 4	74	64
5x8=	89	96	62 x 3	82	76
0 x 5	82	89	53 x 4	70	64
10 x 6	88	91	72 x 2	84	78
32 x 2	82	76	143 x 2	76	71
17 x 3	67	66	113 x 9	63	63

47/C

Site 01

CTBS Multiplication versus Covariate (Kuhlmann-Anderson)

Third Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		7	8
Covariate	Mean	49.8	53.2
	Standard Deviation	7.6	6.8
CTBS: Multiplication	Mean	9.3	9.3
	Standard Deviation	1.3	1.5
	Adjusted Mean	9.5	9.1
Probability Level of Differences		.63	

CTBS Item Analysis, Third Grade, Site 01

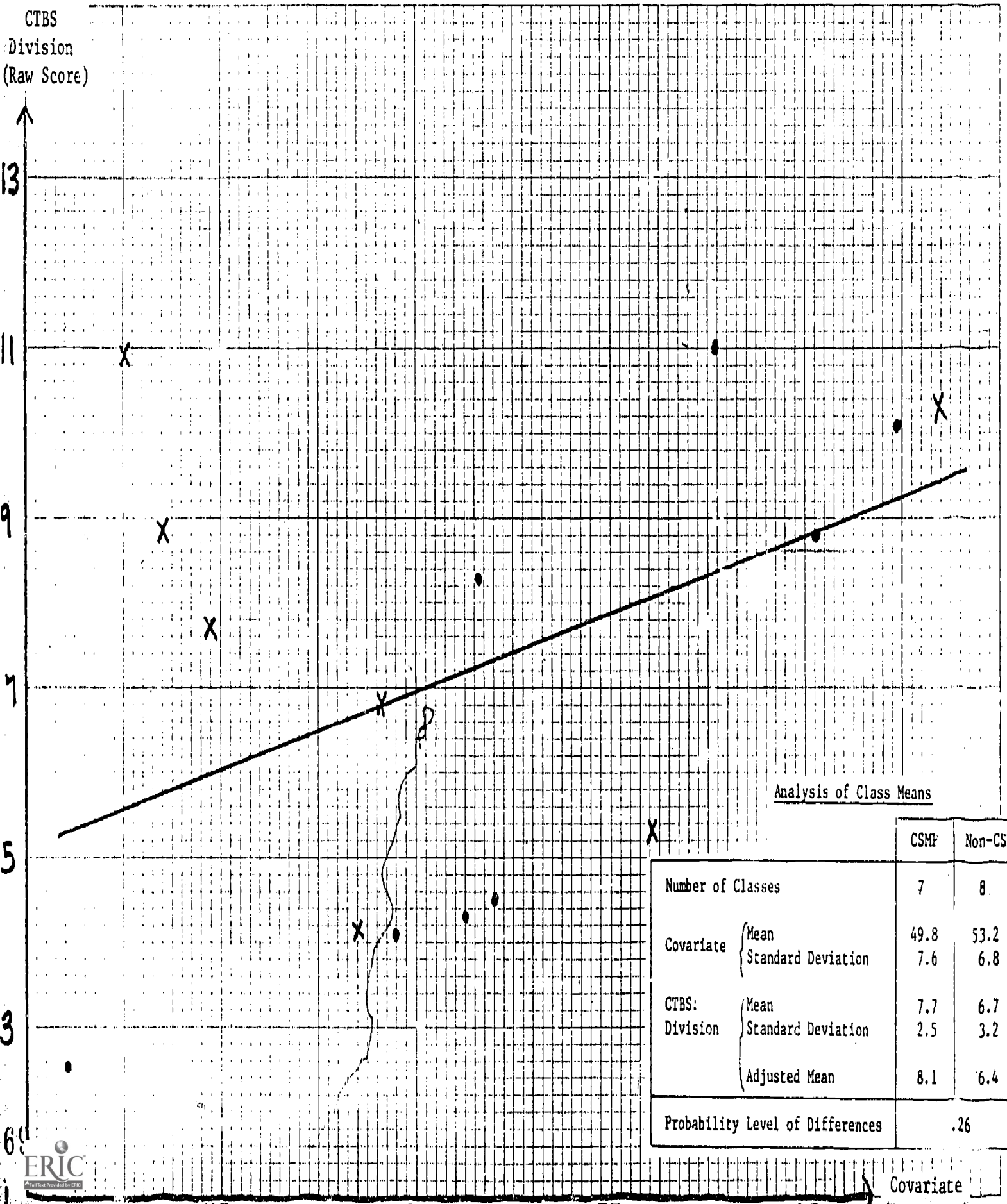
Subtest 4: Division (12 items)

	Percent Correct				Percent Correct		
	CSMP	Non-CSMP			CSMP	Non-CSMP	
$3\overline{)9}$	76	66	*	$8\overline{)24}$	70	61	
$2\overline{)8}$	80	70	*	$6\overline{)36}$	69	65	
$1\overline{)2}$	58	59		$7\overline{)35}$	60	64	
$5\overline{)25}$	79	67	*	$4\overline{)124}$	60	46	*
$3\overline{)15}$	67	63		$24\overline{)48}$	62	39	*
$5\overline{)10}$	70	62		$90\overline{)90}$	40	37	

Site 01

CTBS Division versus Covariate (Kuhlmann-Anderson)

Third Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		7	8
Covariate	Mean	49.8	53.2
	Standard Deviation	7.6	6.8
CTBS: Division	Mean	7.7	6.7
	Standard Deviation	2.5	3.2
	Adjusted Mean	8.1	6.4
Probability Level of Differences		.26	

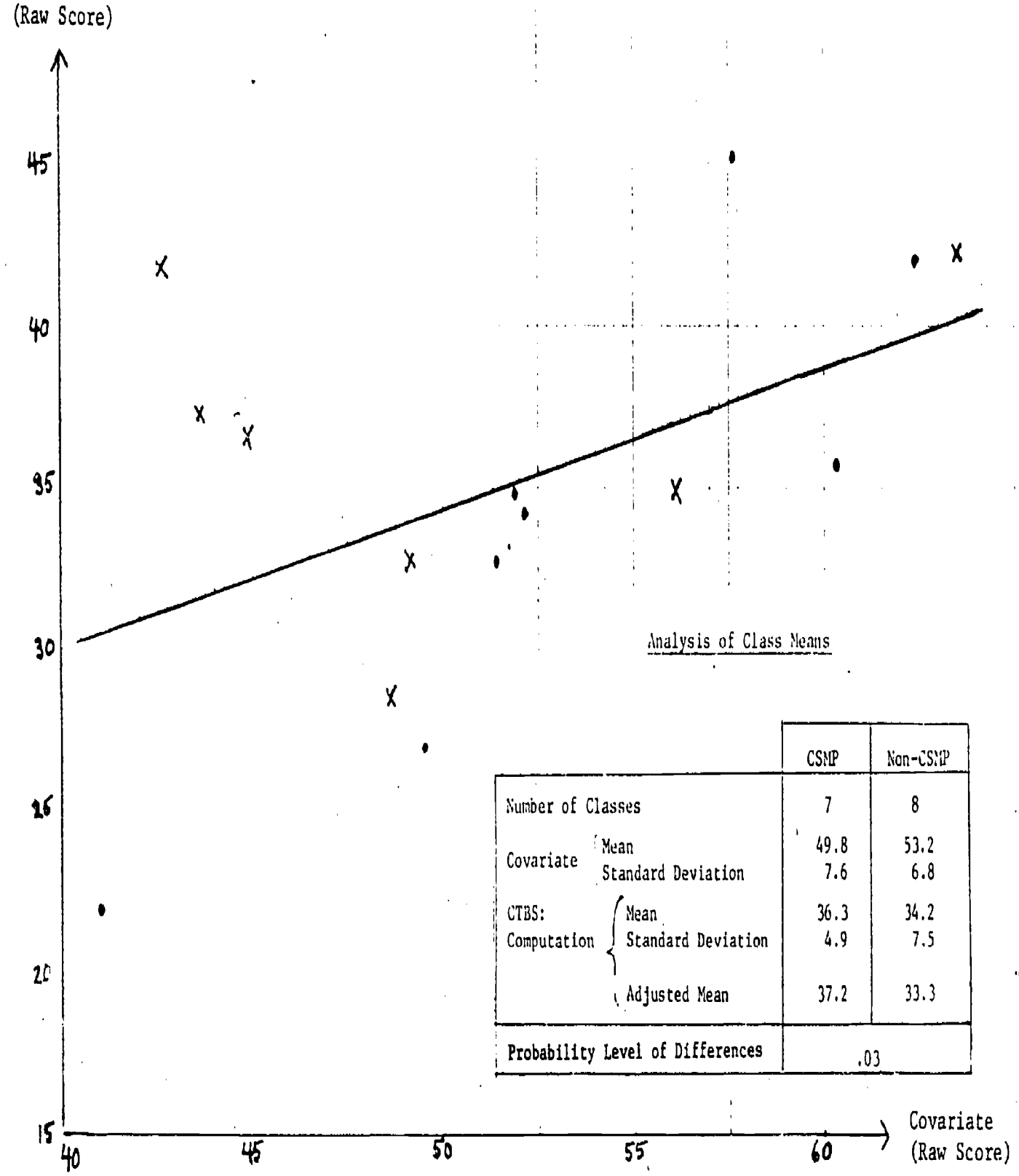
CTBS, Third Grade, Site 01

The graph on the next page is for the Computation Test which is composed of the four previous sets of 12 items each for addition, subtraction, multiplication and division.

Site 01

CTBS Computation versus Covariate (Kuhlmann Anderson)

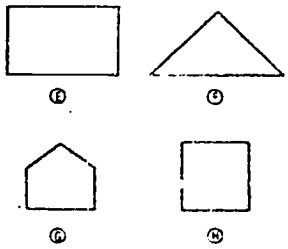
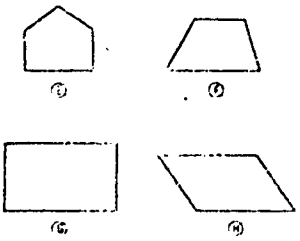
Third Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)

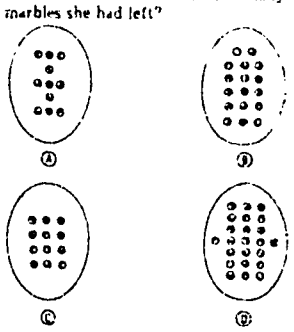
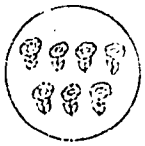
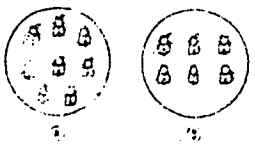
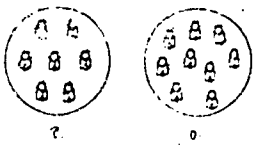
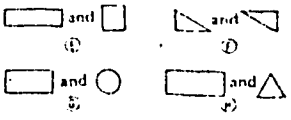


Analysis of Class Means

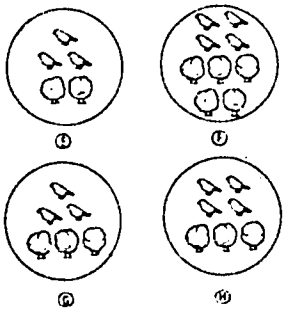
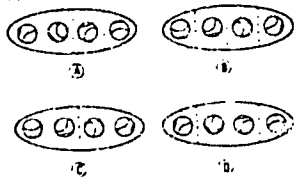
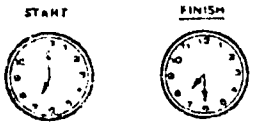
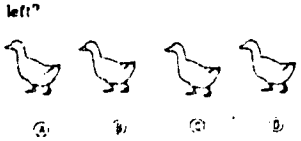
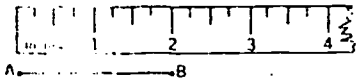
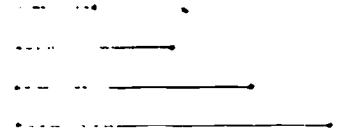
		CSMP	Non-CSMP
Number of Classes		7	8
Covariate	Mean	49.8	53.2
	Standard Deviation	7.6	6.8
CTBS: Computation	Mean	36.3	34.2
	Standard Deviation	4.9	7.5
	Adjusted Mean	37.2	33.3
Probability Level of Differences		.03	

Subtest 5: Concepts and Applications (50 items)

		Percent Correct	
		CSMP	Non-CSMP
1	What does the "3" in 13 stand for? Ⓐ 3 ones Ⓑ 13 ones Ⓒ 3 tens Ⓓ 13 tens	46	42
2	Which of these figures is a square?  Ⓕ Ⓖ Ⓗ Ⓘ	86	88
3	Which of these numerals is twenty-three? Ⓐ 0.23 Ⓑ 2.3 Ⓒ 23 Ⓓ 33	84	84
4	Which of these figures is a rectangle?  Ⓕ Ⓖ Ⓗ Ⓘ	76	71
5	Which of these numerals is two hundred six? Ⓐ 26 Ⓑ 206 Ⓒ 260 Ⓓ 2,000	35	92
6	Look at the circle.  Which letter shows the center of the circle? Ⓓ E Ⓔ F Ⓕ G Ⓖ H	84	81
7	Which of these numerals shows three hundreds, four tens, and six ones? Ⓐ 346 Ⓑ 364 Ⓒ 163 Ⓓ 643	87	89
8	Look at the circle.  Which letter shows a point on the circle? Ⓓ E Ⓔ F Ⓕ G Ⓖ H	67	65
9	Which of these is most like a circle? Ⓐ book Ⓑ chair Ⓒ wheel Ⓓ door	90	5[illegible]

		Percent Correct	
		CSMP	Non-CSMP
27	Nina had a set of 17 marbles. She lost 6 marbles in a game. Which one of the sets below shows how many marbles she had left?  Ⓐ Ⓑ Ⓒ Ⓓ	82	79
28	Here are a set of rabbits and a set of carrots. How many more carrots are there than rabbits?  Ⓔ 3 Ⓕ 4 Ⓖ 7 Ⓗ 11	82	74
29	Here is a set of flowers  Which one of the sets of bees below has fewer members than the set of flowers?  Ⓔ Ⓕ  Ⓖ Ⓗ	58	50
30	Which one of these pairs of figures could be put together to make a rectangle?  Ⓒ Ⓓ	69	64
31	Fran wants to buy a ticket for 60¢. Which of the following shows how she could pay for it? Ⓐ five nickels Ⓑ two dimes and a nickel Ⓒ six dimes Ⓓ three [illegible]	85	83

Subtest 5: Concepts and Applications (cont.)

		Percent Correct	
		CSMP	Non-CSMP
10	Which of the sets below has an equal number of birds and trees? 	83	76
11	Which of the sets below is separated into fourths? 	80	84
12	How many members are there in an empty set? Ⓕ 0 Ⓖ 1 Ⓗ 2 Ⓘ 10	78	81
13	Mr. Smith washed his car. The two clocks show you when he started and when he finished. At what time did he finish?  Ⓐ 6:40 Ⓑ 7:30 Ⓒ 8:00 Ⓓ 8:30	60	67
14	What do centimeter, meter and kilometer all measure? Ⓕ time Ⓖ money Ⓗ weight Ⓘ distance	47	44
15	A dollar has the same value as Ⓐ 5 nickels Ⓑ 10 nickels Ⓒ 20 nickels Ⓓ 25 nickels	54	46
16	How many eggs are there in one-half dozen? Ⓕ 2 Ⓖ 6 Ⓗ 12 Ⓘ 24	71	62
17	Which duck is third in line from the left? 	77	76
18	Which of these is another name for 14? Ⓕ 19 - 4 Ⓖ 16 - 3 Ⓗ 6 + 7 Ⓘ 7 + 7	93	88
19	Which of the following makes this number sentence true? 8 + 6 = 6 + □ Ⓐ 6 Ⓑ 8 Ⓒ 12 Ⓓ 14	57	* 39
32	Which of these shows how much money a ten-dollar bill and one cent equal? Ⓕ $10.01 Ⓖ $101 Ⓗ $10.10 Ⓘ $110	74	66
33	What is the date one week after April 8? Ⓕ April 9 Ⓖ April 14 Ⓗ April 15 Ⓘ April 18	35	38
34	Larry had a rope 2 feet long. He cut it in half. How long was each half? Ⓐ 6 inches Ⓑ 12 inches Ⓒ 24 inches Ⓓ 48 inches	36	43
35	If 4 + n = 12, then n is Ⓐ 3 Ⓑ 4 Ⓒ 6 Ⓓ 12	73	66
36	Which line segment below is one inch longer than segment AB?  	47	38
37	What is the missing number in 3, 5, 7, ____, 11? Ⓕ 1 Ⓖ 8 Ⓗ 9 Ⓘ 10	74	* 64
38	What goes in the box to make this number sentence true? □ + 7 = 4 + 8 Ⓕ 1 Ⓖ 6 Ⓗ 12 Ⓘ 19	66	* 54
39	What goes in the box to make this number sentence true? 5 + 4 = 10 - □ Ⓐ 0 Ⓑ 1 Ⓒ 9 Ⓓ 10	63	* 51
40	x = 3 y = 4 If x + y = z, then z is Ⓕ 5 Ⓖ 9 Ⓗ 10 Ⓘ 20	47	47
41	Joel bought 10 candies. He ate 2 and gave 3 to Pat. How many candies did he have left? Ⓐ 5 Ⓑ 7 Ⓒ 8 Ⓓ 15	71	74
42	Mario had 8 pairs of socks. Half of them were new. How many pairs of socks were old? Ⓕ 2 Ⓖ 4 Ⓗ 8 Ⓘ 16	71	73
43	Sarah had 2 books and Mark had 7 books. They gave 3 books to John. How many books did Sarah and Mark have left all together? Ⓐ 2 Ⓑ 6 Ⓒ 9 Ⓓ 12	61	58
44	Bill is taller than Ted. Juan is taller than Bill. Which of the three boys is the tallest? Ⓕ Juan Ⓖ Ted Ⓗ Bill Ⓘ cannot tell	55	51

CTBS Item Analysis, Third Grade, Site 01

Subtest 5: Concepts and Applications (cont.)

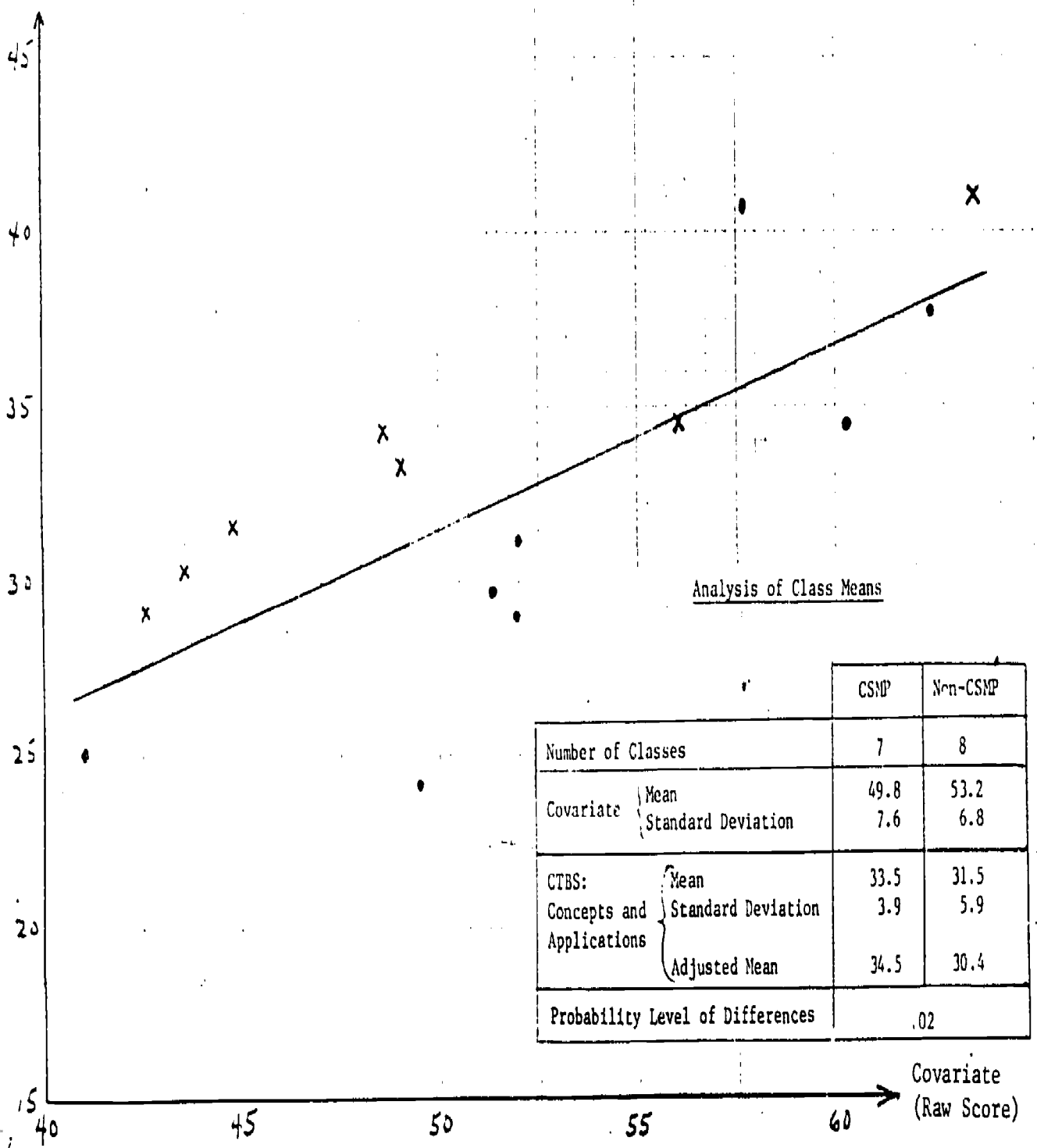
		Percent Correct CSMP	Percent Correct Non-CSMP
20	Which of these is another name for second? Ⓐ 1st Ⓑ 2nd Ⓒ 12th Ⓓ 22nd	88	88
21	Which of these numbers is nearest to 500 on the number line? Ⓐ 496 Ⓑ 498 Ⓒ 503 Ⓓ 504	69	61
22	Which of these numbers is greater than 88,855? Ⓐ 88,855 Ⓑ 88,558 Ⓒ 88,858 Ⓓ 88,585	61	56
23	Which of the following makes this number sentence true? $3 \times (2 \times 4) = (3 \times 2) \times \square$ Ⓐ 4 Ⓑ 6 Ⓒ 8 Ⓓ 24	50	32
24	Which number pair below has an odd number and an even number? Ⓐ (6,3) Ⓑ (8,2) Ⓒ (24,4) Ⓓ (27,3)	46	46
25	Paul had 50¢. He spent 15¢ for milk and found 5¢. What can you do to find out how much money Paul had then? Ⓐ divide and add Ⓑ subtract and add Ⓒ divide and multiply Ⓓ subtract and multiply	58	54
26	Anita picked 4 apples and Louis picked 9 apples. Which one of the sets below shows how many apples they picked all together? Ⓐ Ⓑ Ⓒ Ⓓ	81	85
45	Sandy had 24 red buttons, 33 black buttons, and 56 white buttons. How many buttons did she have in all? A 98 B 103 C 113 D 123	56	53
46	Otis found 23 sea shells. He would like to have 52 sea shells. How many more shells must he find? Ⓐ 23 Ⓑ 29 Ⓒ 52 Ⓓ 75	53	47
47	Ann had 4 records. She got 2 more records for her birthday and gave 1 record to her sister. What can you do to find how many records she had then? Ⓐ add and divide Ⓑ add and subtract Ⓒ multiply and divide Ⓓ multiply and subtract	57	68
48	Maria had 20¢ to buy kite string. It cost 10¢ for 60 yards. How many yards of string could she buy? Ⓐ 20 yards Ⓑ 60 yards Ⓒ 120 yards Ⓓ 200 yards	50	49
49	Charles had 7 peanuts and Pedro had 2 times as many. How many peanuts did Pedro have? A 9 B 14 C 16 D 21	59	43
50	Mr Alvarez had 36 prizes. He divided the prizes equally among 6 children. How many prizes did each child get? E 6 F 8 G 32 H 42	60	56

55

Site 01

CTBS Concepts and Applications versus Covariate (Kuhlmann Anderson)

Third Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		7	8
Covariate	Mean	49.8	53.2
	Standard Deviation	7.6	6.8
CTBS: Concepts and Applications	Mean	33.5	31.5
	Standard Deviation	3.9	5.9
	Adjusted Mean	34.5	30.4
Probability Level of Differences		.02	

Commentary

The difference in mean adjusted scores was about four points in favor of the CSMP classes on both the Computation and the Concepts and Applications Tests. These differences were large enough to be significant, even with the relatively small number of classes involved. However, it is interesting to compare the graphs of these two tests. On the Concepts and Applications Test, the data fit the regression line quite well and the CSMP classes are generally doing better at all levels of ability. On the Computation Test, however, there is much more dispersion about the regression line (in fact for the CSMP classes only, the distribution is more U-shaped than linear); the difference in the two groups seems to derive from the three CSMP classes of lowest ability, all of whom have high scores on the Computation Test (these are the x's in the upper left portion of the graph).

In looking at the percentages correct on various items, it can be seen that CSMP students seem to do particularly well, relative to Non-CSMP students, on items in the Concepts and Applications Test involving open sentences and of the form: "What number goes in the box to make this number sentence true? ...". In the set of addition items from the Computation Test, CSMP students did relatively best on the two items in which the addition problem was given in horizontal format; e.g. 72+73=__ .

Appendix D

CSMP-Specific Tests A and B, Second Grade

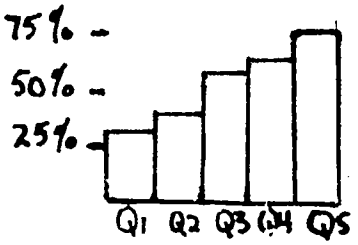
For each of CSMP-Specific Tests A and B, similar pages have been grouped together in the analysis to form subtests. A total of seven subtests were thus formed and the data from these subtests are presented one after the other, in the following order:

CSMP-Test A: Subtest 1, Arrow Diagrams
Subtest 2, Minicomputer
Subtest 3, Integers

CSMP-Test B: Subtest 1, Arrow Diagrams
Subtest 2, Minicomputer
Subtest 3, String Pictures
Subtest 4, Computation

For each subtest, three kinds of information are presented: a) item analysis data, b) summary subtest data and c) class mean data.

a) Item Analysis Data. For each test page, a page appears in the Appendix on which three pieces of information are given:
   i) At the top of the page are any directions for that page that were given by the tester to the whole class before they began that page, where such were considered necessary. It should be noted that these tests contained no new kinds of problems (problems not found in regular workbooks), and that additional directions, sometimes given individually to students, are not given here.
   ii) In the middle of the page is a reduced version of the actual test page. For each item on this test page two kinds of data are given. A circled entry (ⓧⓧ) indicates the percent correct based on all students who took the test (n=650-670) and a bar graph entry shows the percent correct for each of five ability levels (quintiles) of students. These bar graph percentages were based on those students for whom a test score in reading or general ability was available (n=390-410). For example the bar graph below shows about 30% correct for those students in the lowest quintile (i.e. students whose percentile rank on the ability or reading test was less than 20); and about 75% correct for those students in the highest quintile (i.e. students whose percentile rank was 80 or more).



iii) At the bottom of the page is any miscellaneous data for particular test items.

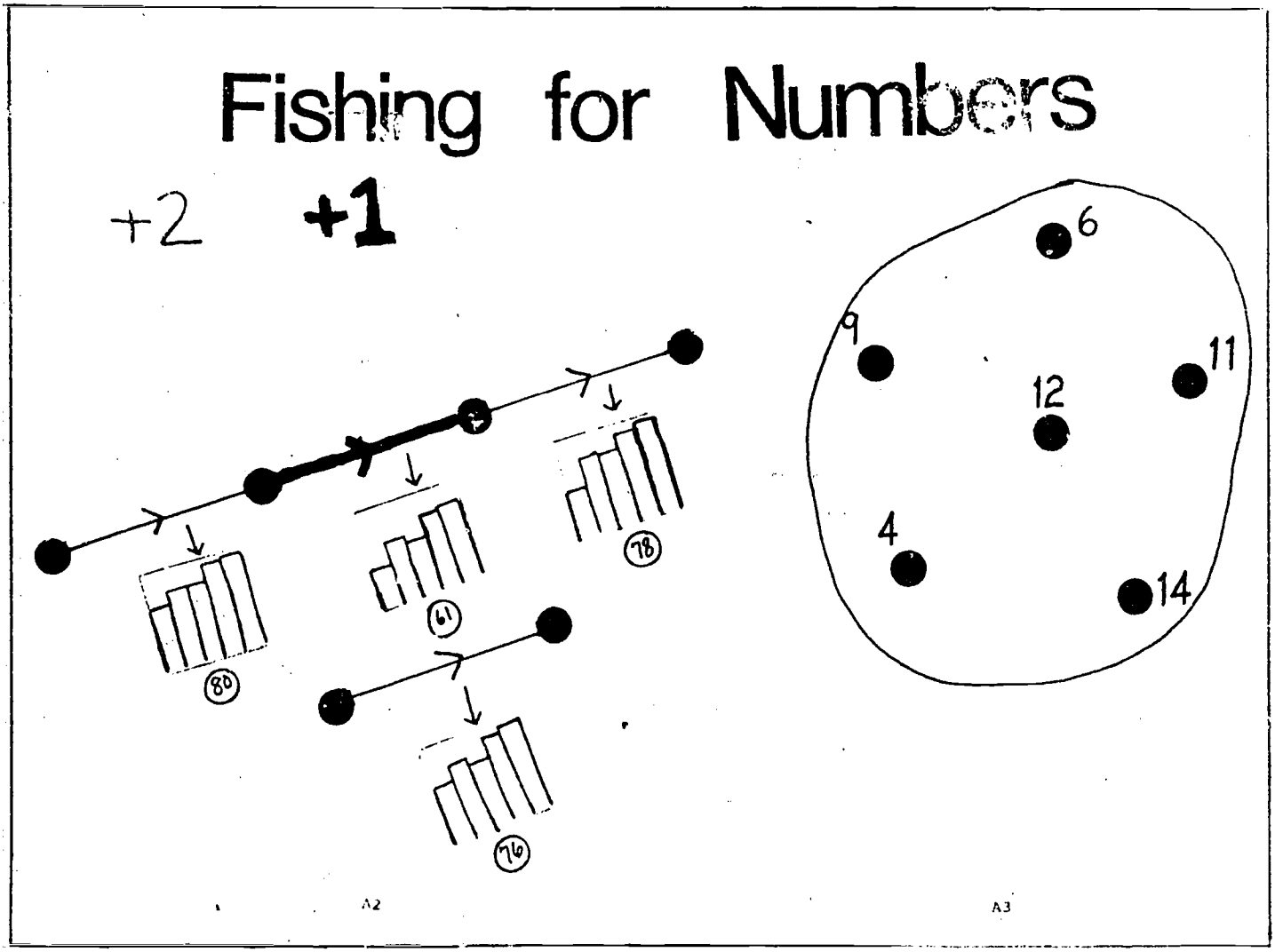
b) Summary Subtest Data. For the subtest (one or more test pages) a frequency distribution and various statistics by quintile according to student ability are given. In particular the percentage of students in each quintile who got all or nearly all the items correct is given as is the percent who got roughly half or fewer of the items correct. Also given are correlation coefficients between the subtest and various other test scores.

c) Class Mean Data. The mean score on the subtest for each class which took the subtest is plotted on a graph against the mean ability or reading score for that class. Through the resulting set of points, one for each class, the regression line has been drawn. The reader can compare these graphs with other subtests (and tests in other appendices) to compare the degree to which class scores are predicted by the covariate (i.e. the degree to which the dots lie close to the regression line).

Finally at the end of this Appendix, summary data, including graphs of class means, are given for the total of CSMP Test A and of CSMP Test B.

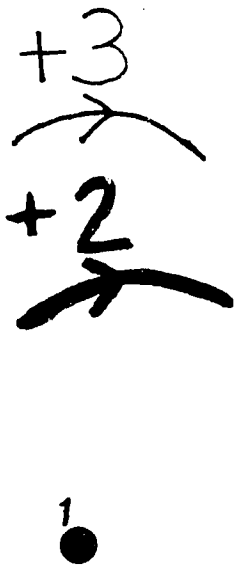
CSMP Test A, Subtest 1, Arrow Diagrams

"The problem is to label each of these dots (point to dots, lower left) with one of these numbers (point to numbered dots within string, lower right). Remember each of the numbers may only be used once. You may have to erase some numbers if it doesn't work out right."



The item statistics are given for each arrow and each was scored independently. An item (arrow) was counted as correct if the dots at the ends of the arrow were labelled so as to "fit" the relation defined by the arrow. For example, the left most blue arrow was counted as correct if the ends were (left to right) 4 and 6, or 9 and 11, or 12 and 14. Fifty-six percent of the students were able to label the 6 dots in the one way that "fit" the arrows, namely 9, 11, 12, 14 respectively along the top set of arrows and 4, 6 for the lower arrow.

CSMP Test A, Subtest 1 Arrow Diagrams (cont'd)



Build a road between 1 and 8.
Use just these arrows.

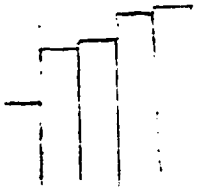
A6

This page was scored on three dimensions, independent of one another:

a) Use of only the given arrows (+3, +2)

(88)

b) Construction of a road which did end at 8.



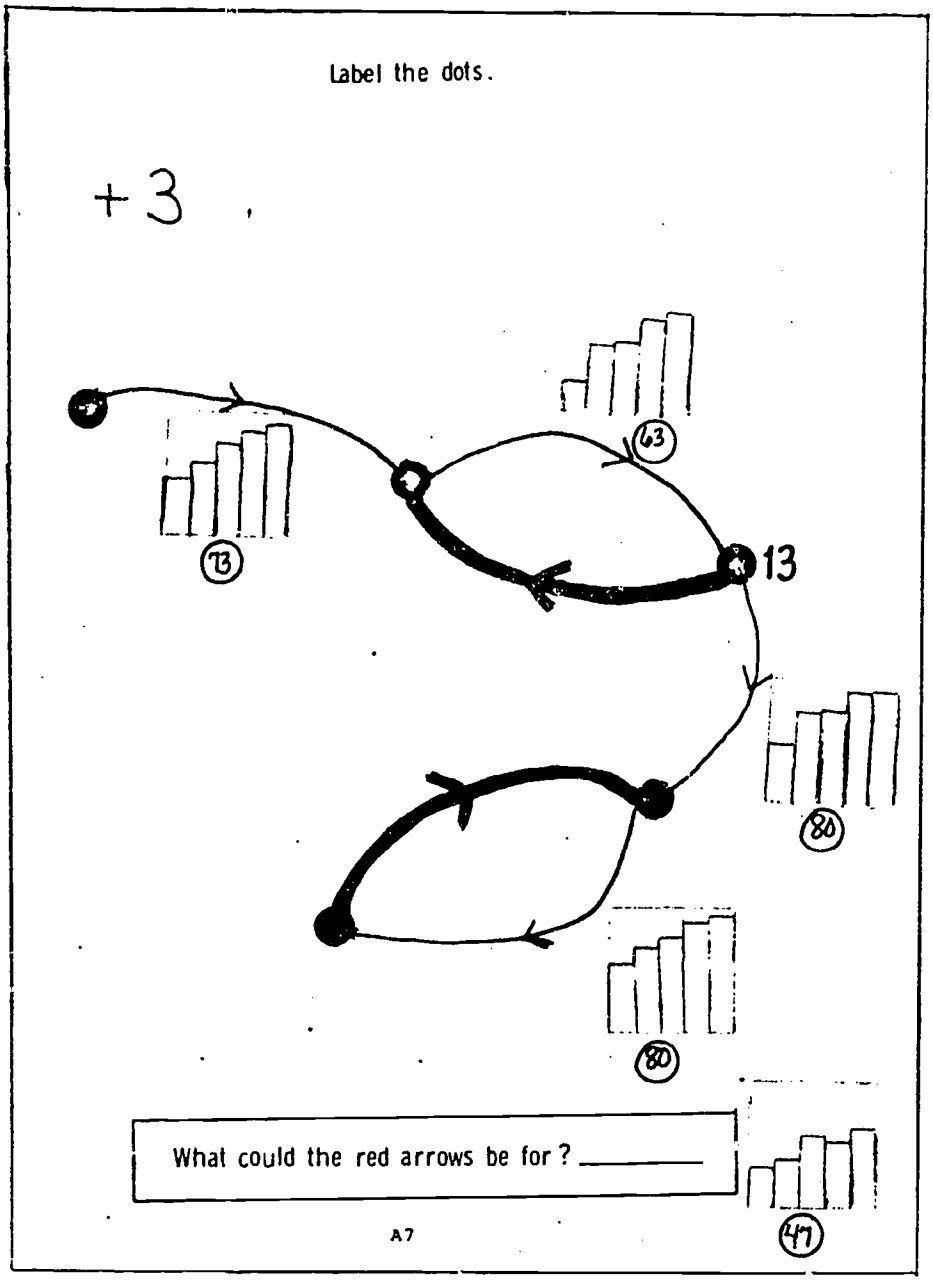
(80)

c) Computational accuracy of all arrows.

(58)

Altogether 58% of the students were able to draw a road from 1 to 8 with the given arrows (i.e. were correct on all 3 dimensions).

CSMP Test A, Subtest 1, Arrow Diagrams (cont'd)

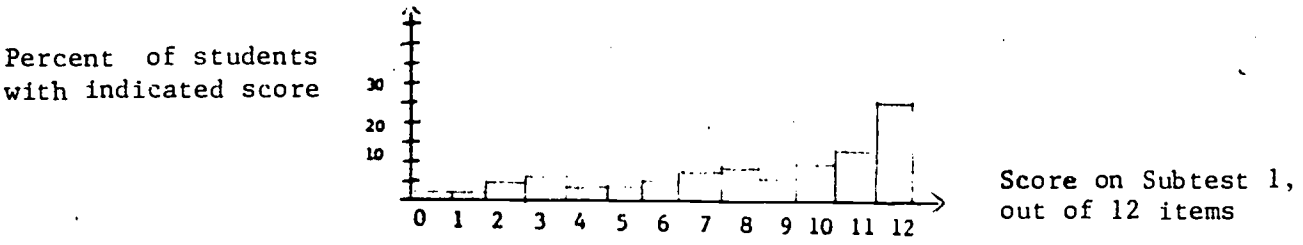


Each arrow was scored independently and was considered correct if the dots at the end of the arrow were labelled so as to "fit" the relation defined by the arrow. The heavy arrows, above, were colored red on the test.

Twenty-one percent of the students omitted the last question ("What could the red arrows be for?") and this number included significant numbers of students in each quintile. Wrong answers to this question tended to be quite different; no single wrong answer was given by many students.

CSMP Test A, Subtest 1: Arrow Diagrams

Frequency Distribution: All Students Combined



Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 1	Standard Deviation	Percent of students with score 0-6	Percent of students with score 11 or 12
Q1	5.7	3.6	55	9
Q2	8.1	3.6	30	38
Q3	8.3	3.3	26	34
Q4	10.1	2.5	12	63
Q5	10.7	1.9	7	72
All Students Combined	8.6	3.4	25	42

[1]Q1 means the lowest quintile, i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

Correlations Between Subtest 1 and Other Tests

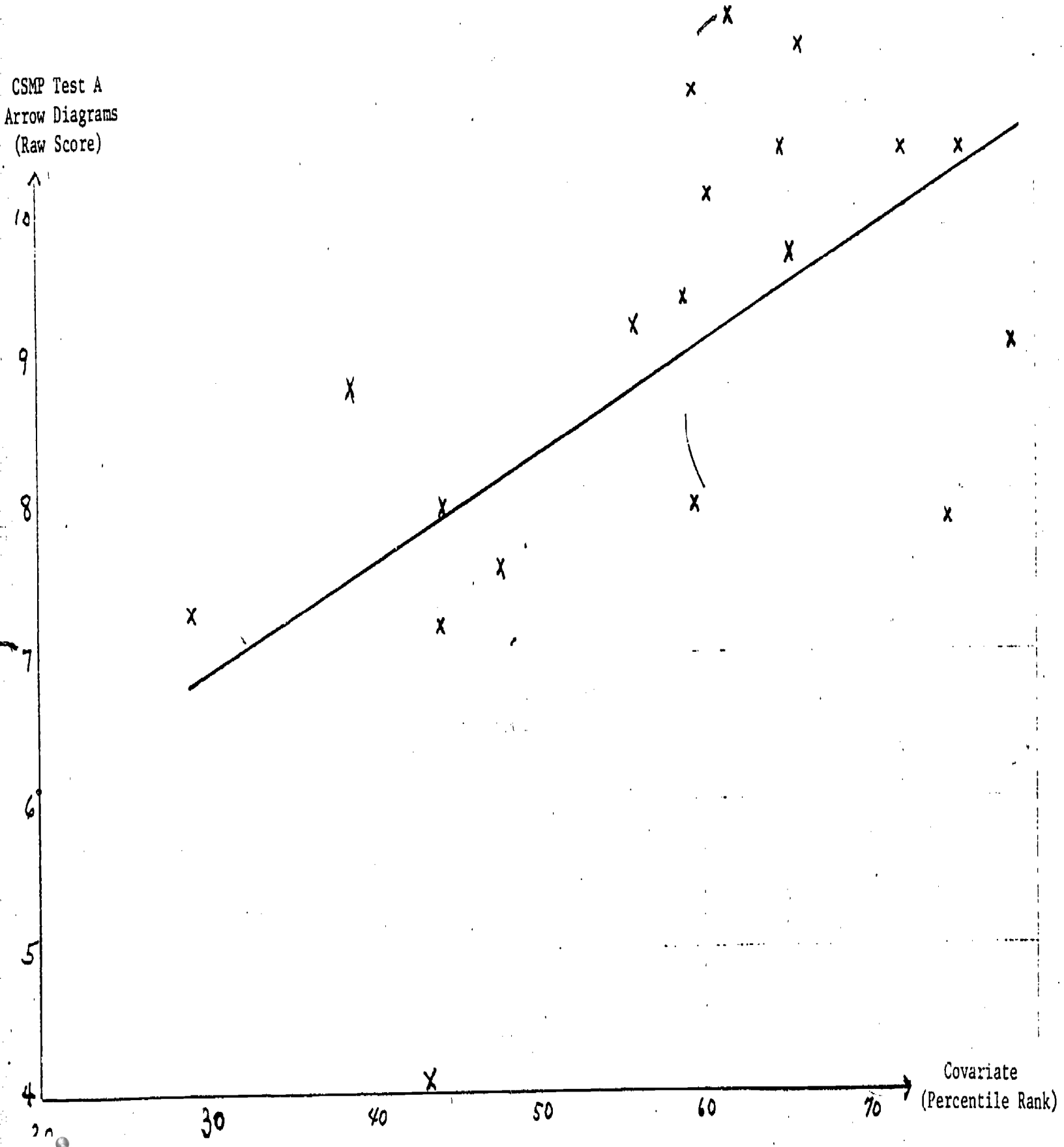
Correlations with (number of students):

Kuhlmann-Anderson:	.67 (124)
SAT Reading:	.50 (83)
CPT Reading:	.33 (95)
Gates McGinitie:	.44 (75)
MANS, Test A:	.66 (294)
MANS, Test B:	.79 (142)
CTBS Math:	.79 (263)

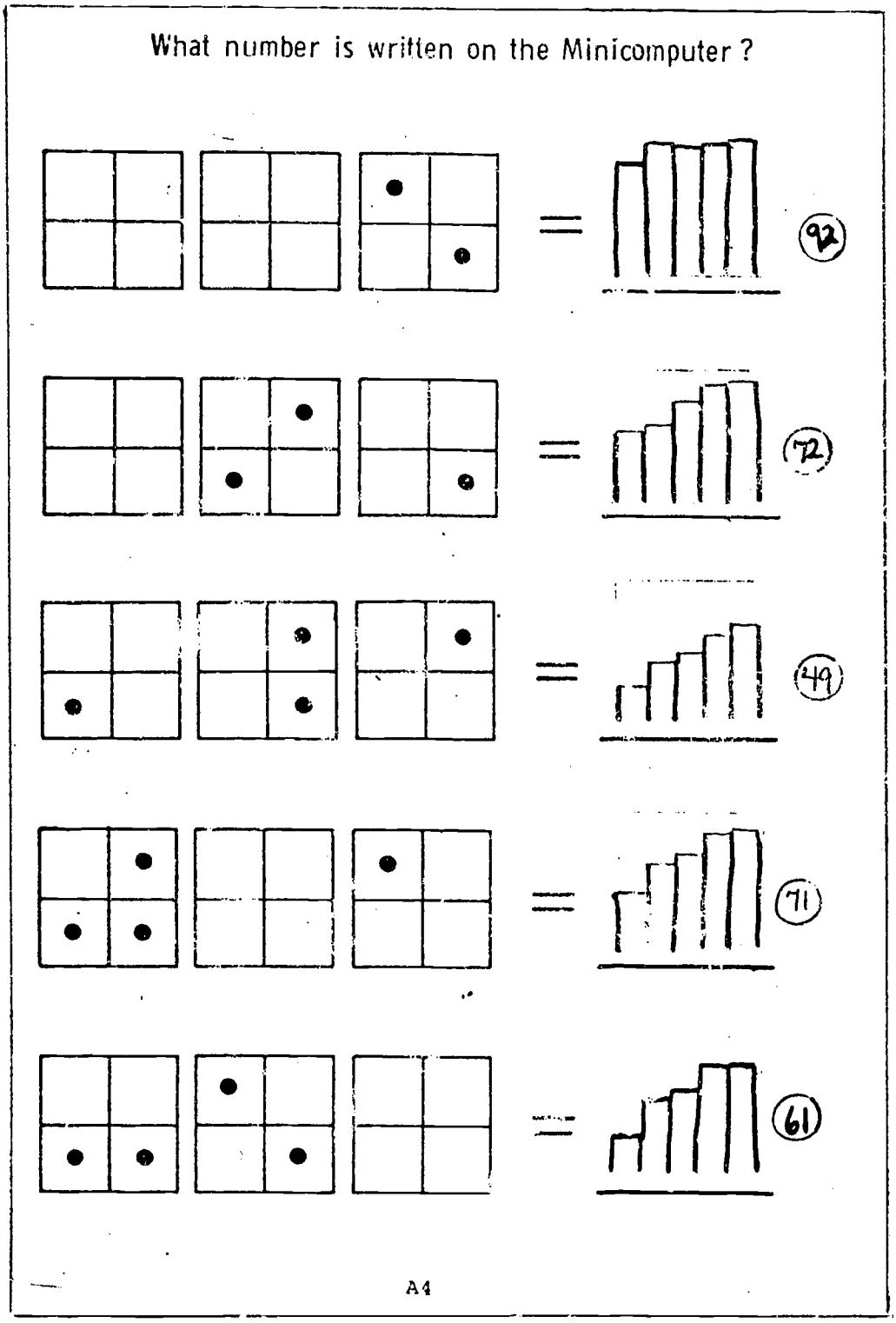
Reliability/Homogeneity (KR20)=.87(652)

63 D

CSMP Test A: Arrow Diagrams Versus Covariate
Second Grade Class Means



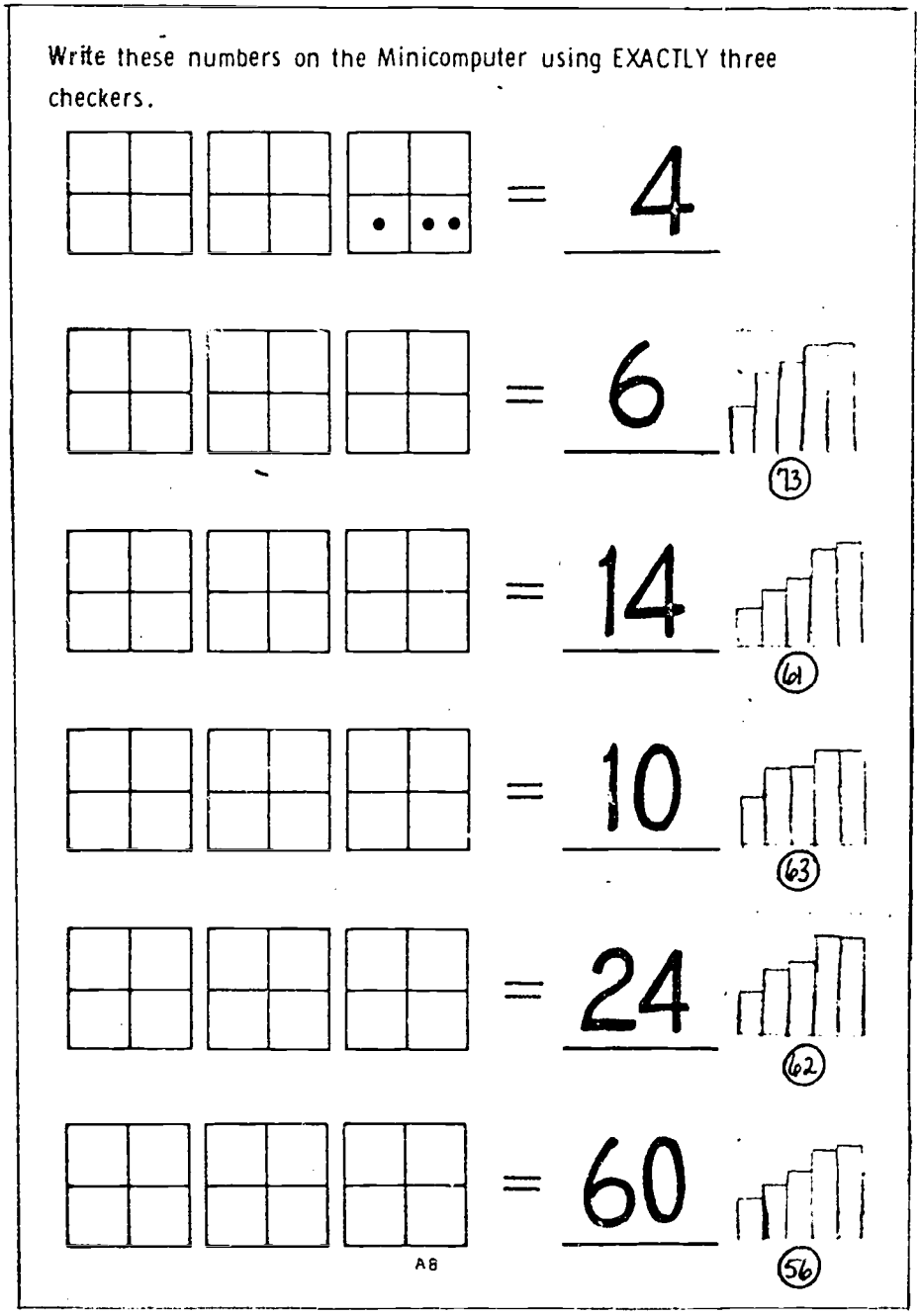
CSMP Test A, Subtest 2, Minicomputer



Fewer than 2% of the students omitted even one item. On the third item, 29% of the students misread one of the three boards (a very common wrong answer being "154"), and on the fourth item, 12% gave a response of either 780 or 78 (instead of 708).

CSMP Test A, Subtest 2, Minicomputer (cont'd)

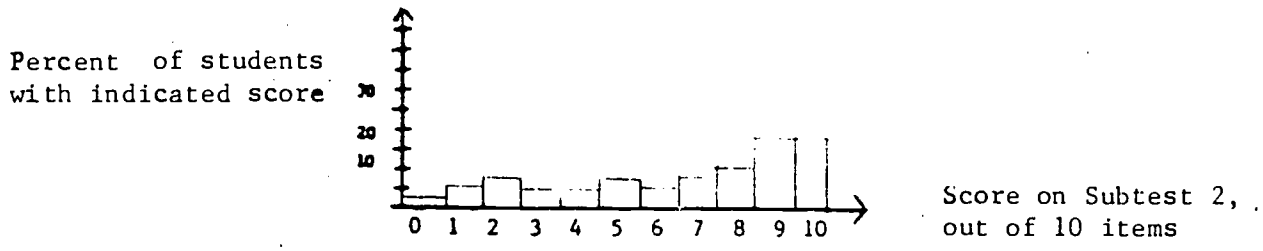
"It says, 'Write the numbers on the Minicomputer using exactly 3 checkers.' Remember you have to use exactly 3 checkers for each one, even if it's easier to do it with 4 checkers or 2 checkers."



An average of 9% of the students showed the required number on the Minicomputer, but with either fewer or more than 3 checkers. An average of 14% of the students used exactly 3 checkers but did not display the required number.

CSMP Test A, Subtest 2: Minicomputer

Frequency Distribution: All Students Combined



Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 2	Standard Deviation	Percent of students with score 0-4	Percent of students with score 9 or 10
Q1	4.0	3.3	67	16
Q2	5.8	3.2	36	30
Q3	6.5	2.7	24	32
Q4	8.0	2.3	14	60
Q5	8.2	2.3	9	60
All Students Combined	6.6	3.0	27	38

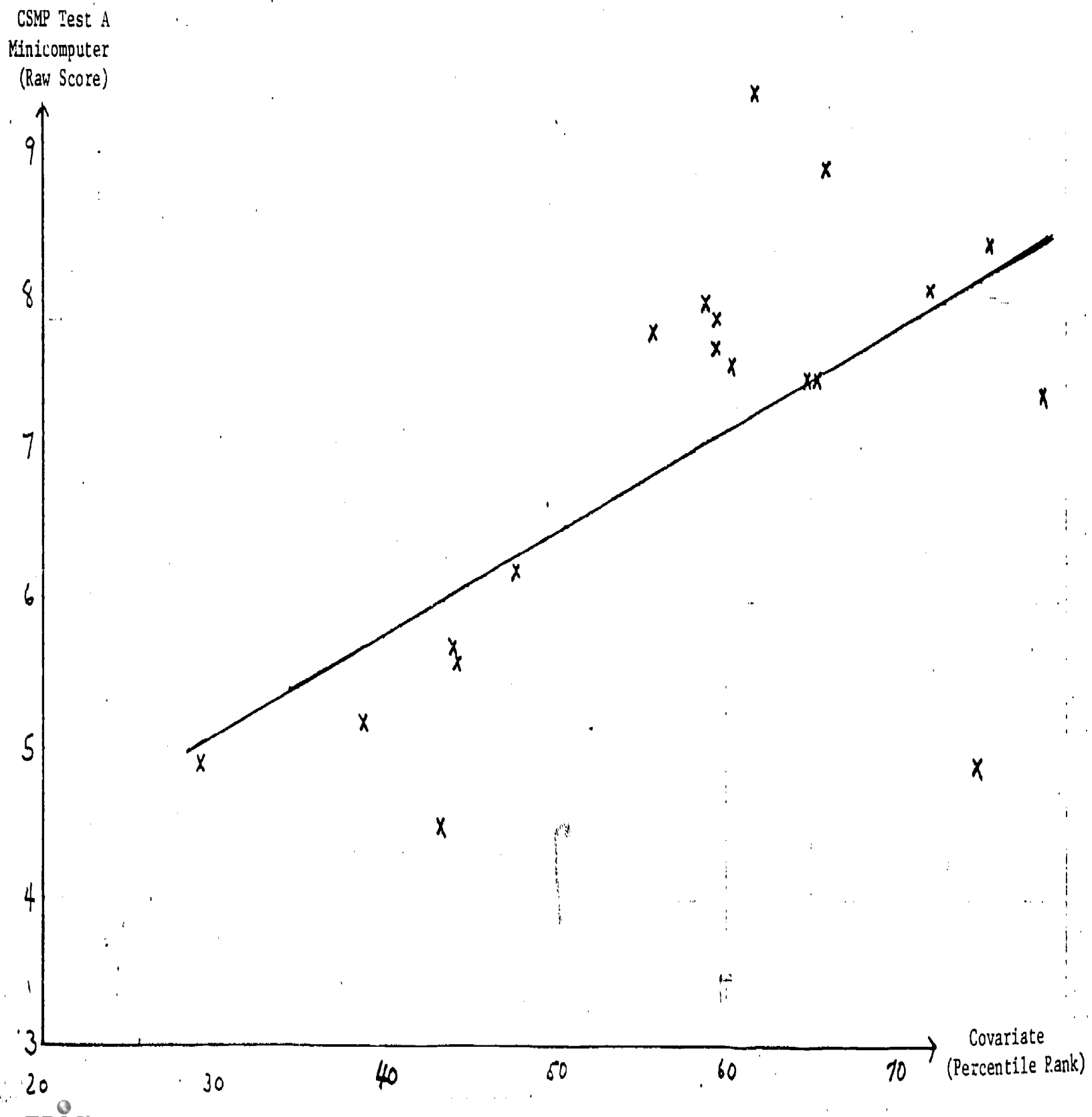
[1]Q1 means the lowest quintile i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

Correlations Between Subtest 2 and Other Tests

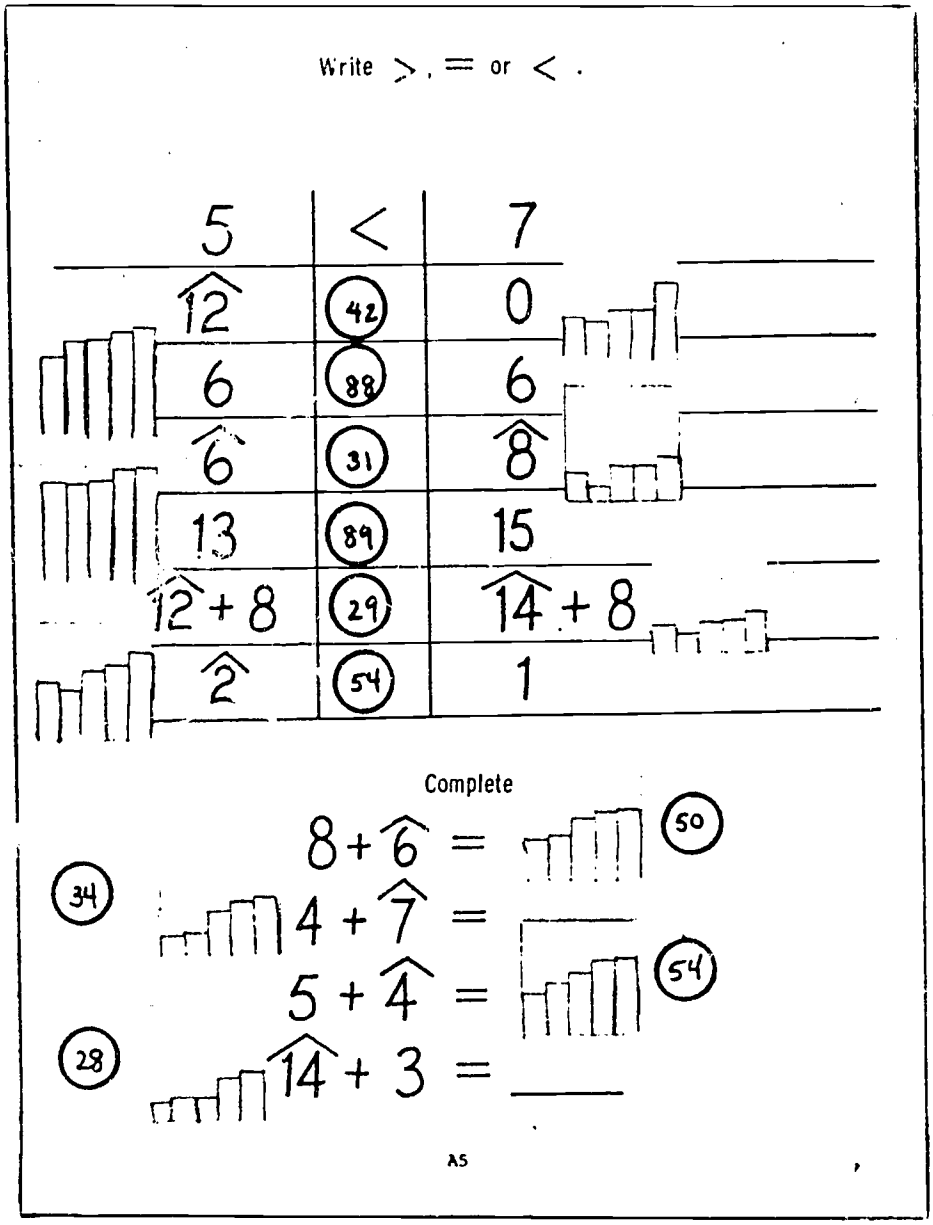
Correlation with (number of students):		
	Kuhlmann-Anderson:	.64 (124)
	SAT Reading:	.37 (83)
	CPT Reading:	.23 (95)
	Gates McGinitie:	.48 (75)
	MANS, Test A:	.69 (294)
	MANS, Test B:	.76 (142)
	CTBS Math:	.77 (263)

Reliability/Homogeneity (KR20)=.86(652)

CSMP Test A: Minicomputer Versus Covariate
Second Grade Class Means



CSMP Test A, Subtest 3, Integers

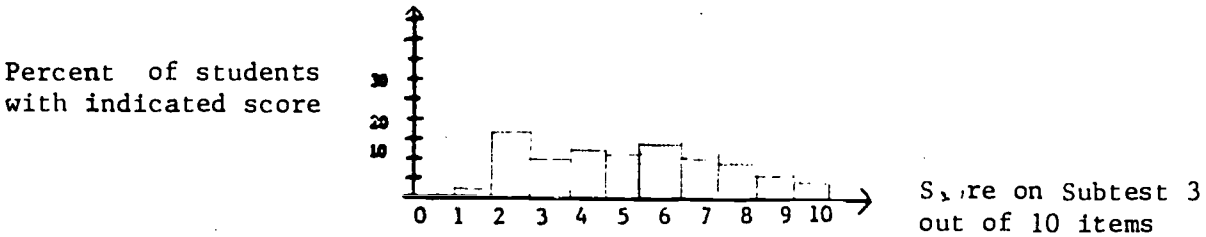


In the top set of items: for the two items without "hats" (^), an average of 89% were correct; for the two items with a negative number on only one side, an average of 48% were correct; and for the two items with negative numbers on each side, an average of 30% were correct. Alternatively, 25% of the students systematically reversed the inequality sign on all 4 items involving negative numbers and a further 14% did the same for only the two items with negative numbers on both sides.

In the bottom set of items: an average of 18% of the students did the computations as if there were no "hats" (gave responses of -14, 11, 9, 17 respectively) and 13% did this systematically on all 4 items; an average of 14% of the students gave the correct absolute value but omitted the hats for the second and fourth items (i.e. the two items in which the given negative number was larger in absolute value than the given positive number), while negligible percentages of students did the reverse - added hats where they shouldn't have been - on the first and third items.

CSMP Test A, Subtest 3: Integers

Frequency Distribution: All Students Combined



Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 3	Standard Deviation	Percent of students with score 0-5	Percent of students with score 9 or 10
Q1	3.9	1.9	78	2
Q2	3.9	2.1	73	2
Q3	5.0	2.4	61	10
Q4	5.6	2.4	42	10
Q5	6.3	2.3	37	19
All Students Combined	5.0	2.6	57	10

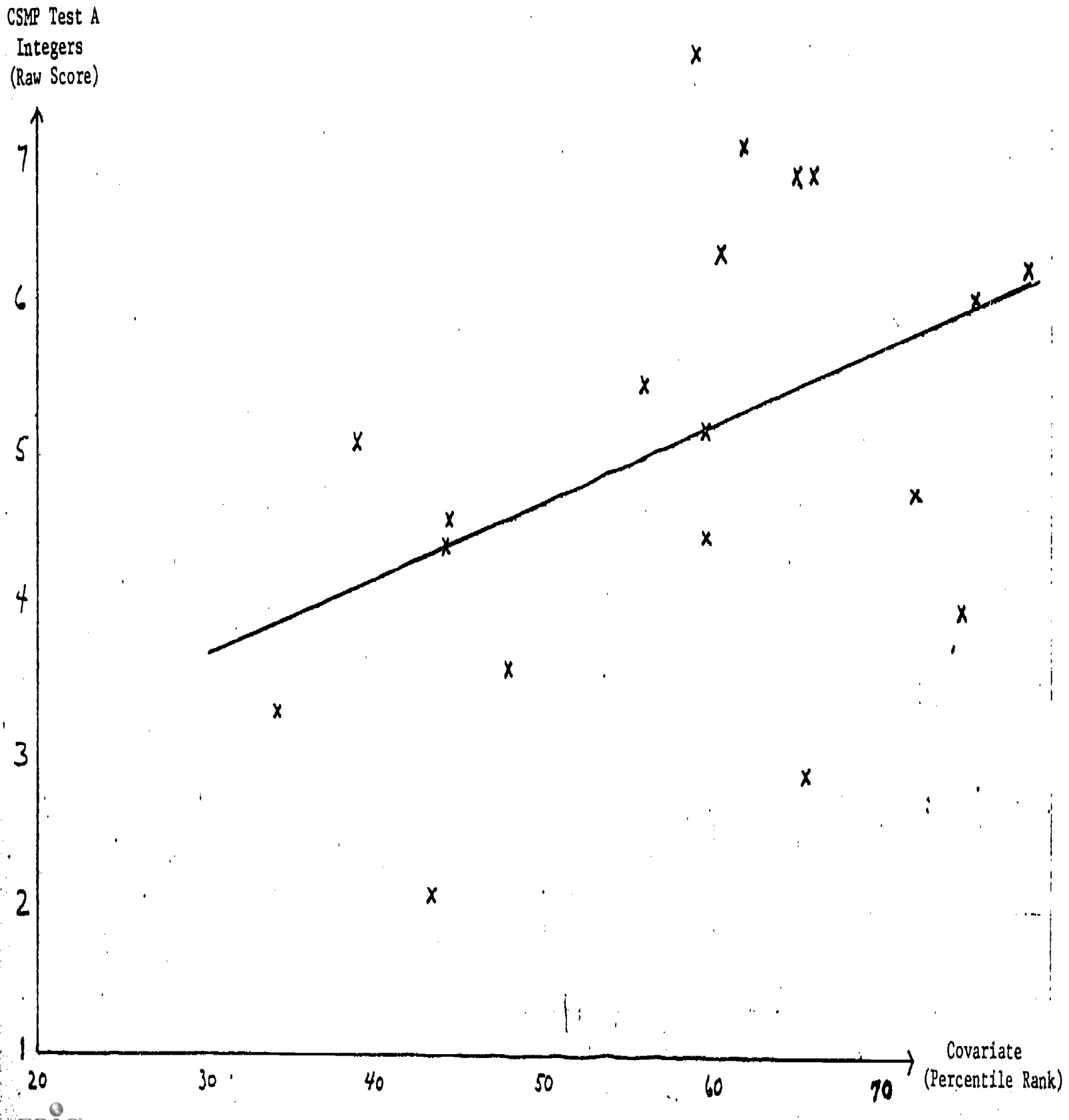
[1]Q1 means the lowest quintile, i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

Correlations Between Subtest 3 and Other Tests

Correlations with (number of students):		
	Kuhlmann-Anderson:	.39 (124)
	SAT Reading:	.34 (83)
	CPT Reading:	.17 (95)
	Gates McGinitie:	.47 (75)
	MANS, Test A:	.57 (294)
	MANS, Test B:	.54 (142)
	CTBS Math:	.54 (263)

Reliability/Homogeneity (KR20)=.76(652)

CSMP Test A: Integers Versus Covariate
Second Grade Class Means



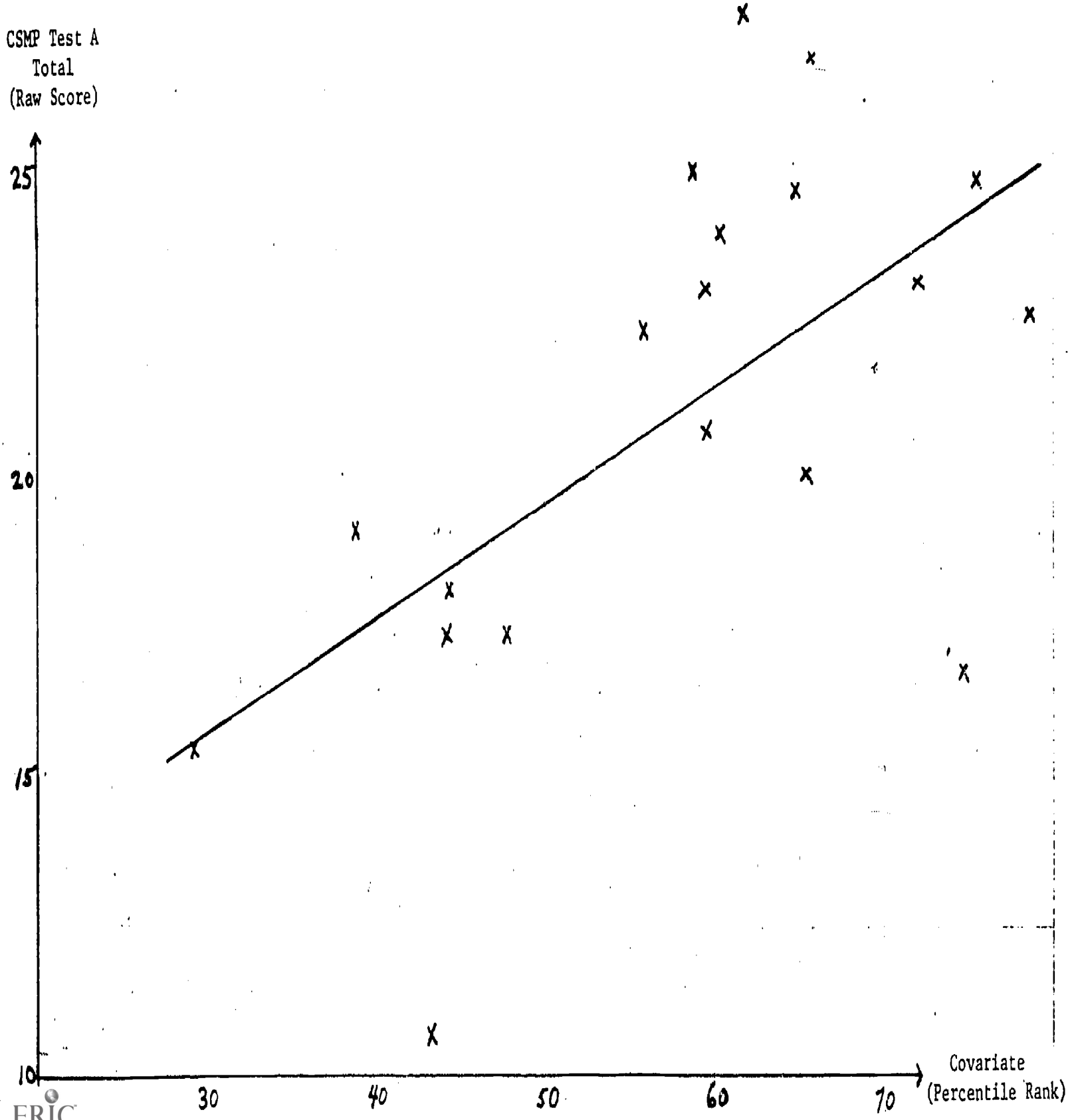
CSMP-Specific Test A, Intercorrelations (n=680)

	Subtest 1	Subtest 2
Subtest 1: Arrow Diagrams		
Subtest 2: Minicomputer	.68	
Subtest 3: Integers	.49	.48

Correlations Between Total Score, Test A and Other Tests

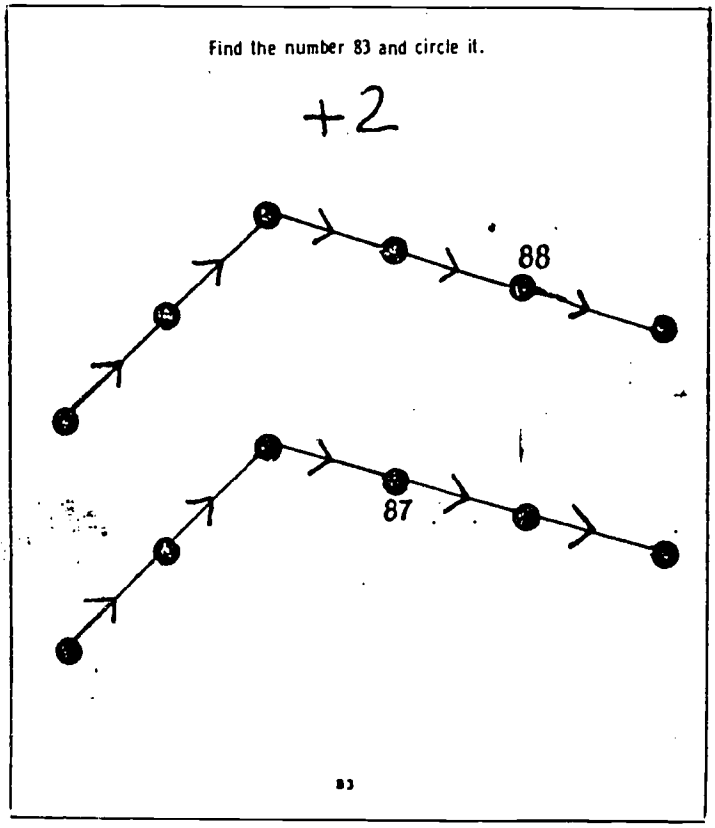
Correlations with (number of students):		
	Kuhlmann-Anderson	.69 (124)
	SAT Reading	.52 (83)
	CPT Reading	.35 (95)
	Gates McGinitie	.53 (75)
	MANS, Test A	.77 (294)
	MANS, Test B	.86 (142)
	CTBS Math:	.82 (263)

CSMP Test A: Total Score Versus Covariate
Second Grade Class Means



CSMP Test B, Subtest 1, Arrow Diagrams

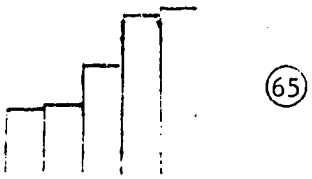
"It says 'Find the number 83 and circle it.' That means that one of the dots on this page is for 83. You have to figure out which one it is and then circle it. You can label any dots you want to help you out."



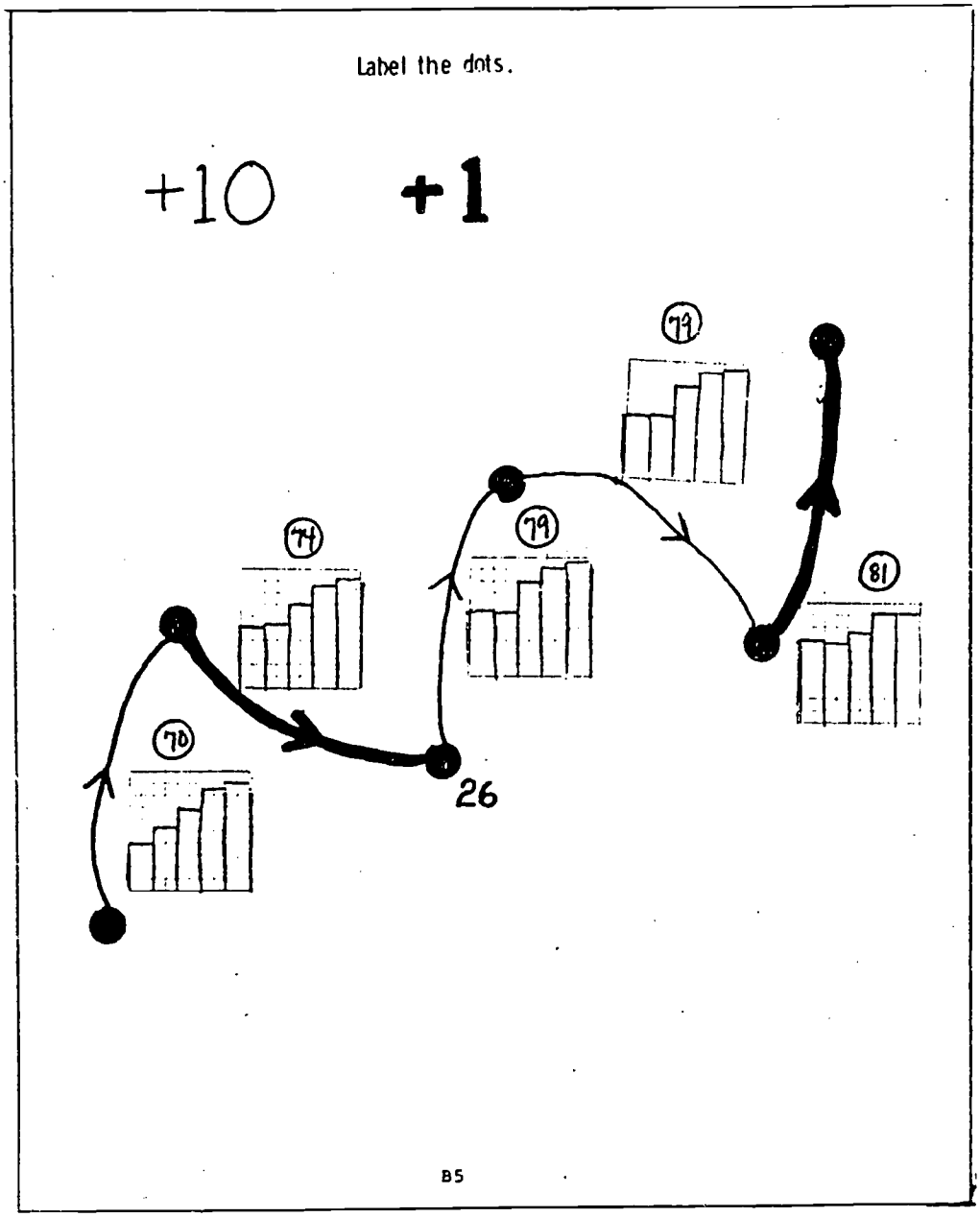
This page was scored on two criteria:

a) Computational accuracy; no dots labelled incorrectly according to the +2 arrows.

b) Circling (or labelling) the correct dot for 83.



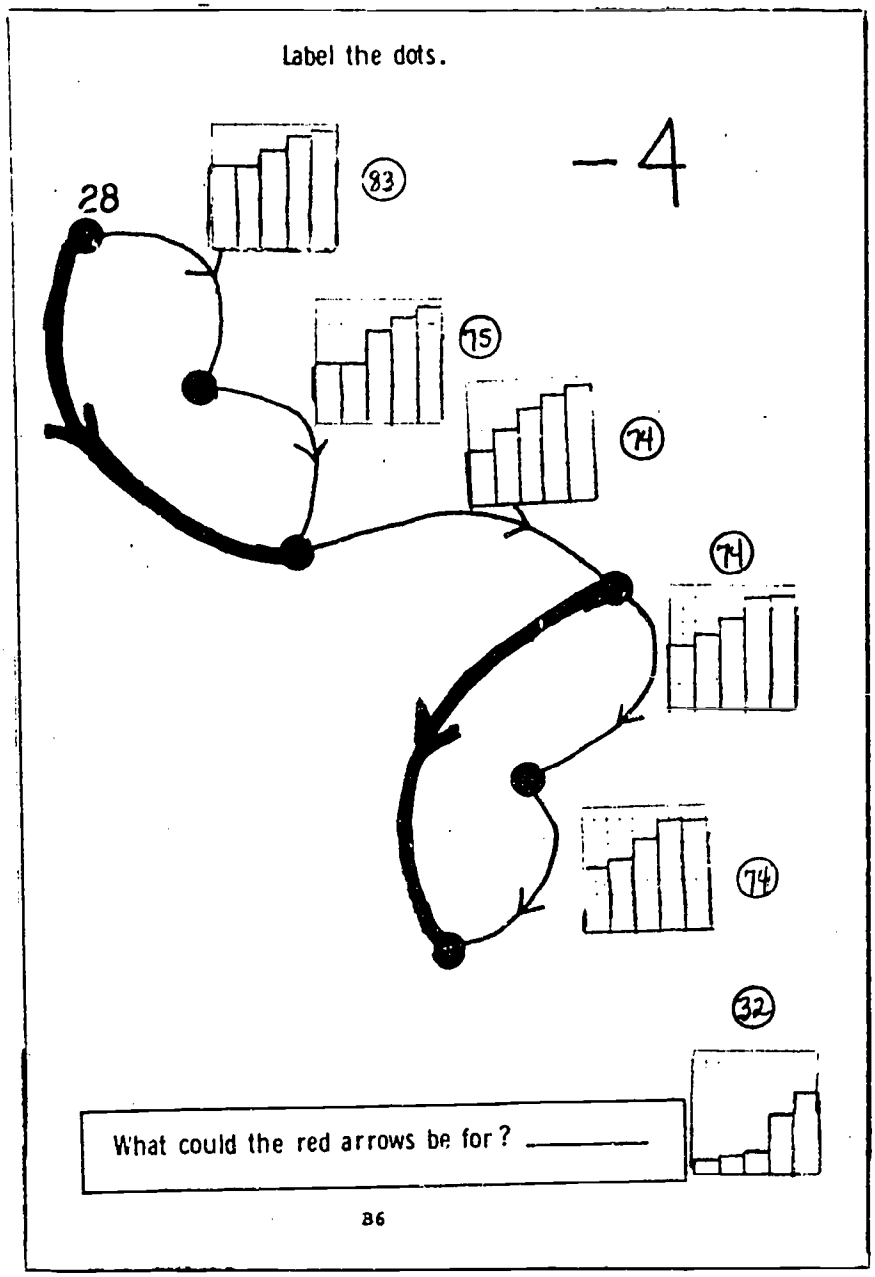
CSMP Test B, Subtest 1, Arrow Diagrams (cont'd)



Each arrow was scored independently and was considered correct if the dots at the end of the arrow were labelled so as to "fit" the relation defined by the arrow.

Fifty-two percent of the students labelled all dots correctly. Averaged across items; only 2% of the students reversed the direction of the arrow and 5% of the students used the wrong arrow (i.e. used the +10 relation instead of the +1 relation, or vice versa).

CSMP Test B, Subtest 1, Arrow Diagrams (cont'd)

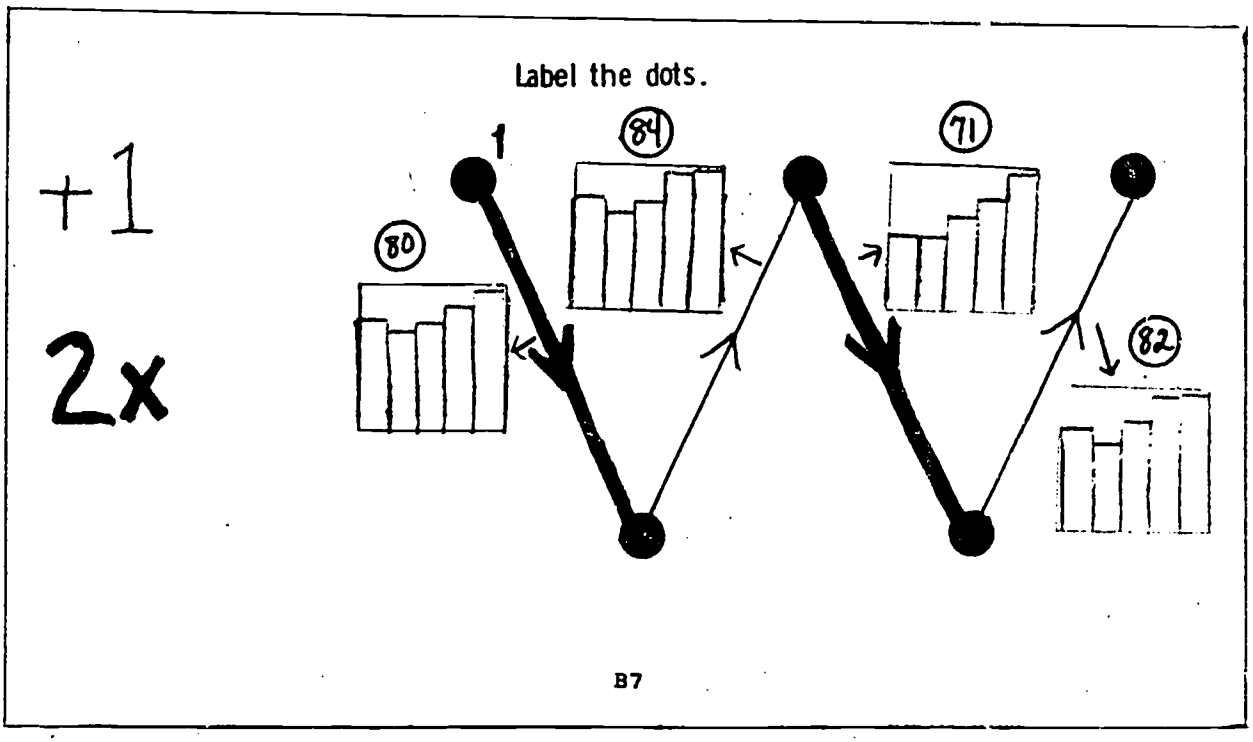


Again, each arrow was scored independently. The heavy arrows were originally red.

The last question, "What could the red arrows be for?", was omitted by 16% of the students. The most frequent answers besides "-8" were "+4", "+8" and "8" (with neither a plus nor a minus sign).

In labelling the dots, an average of 5% of the responses contained a computational error of ±1 and an average of only 2% of the responses used the relation "+4" instead of "-4".

CSMP Test B, Subtest 1, Arrow Diagrams (cont'd)



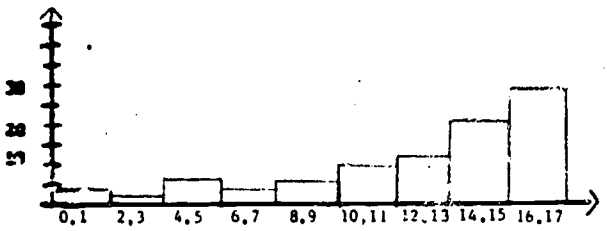
An average of 4% of the responses per item evidently used the wrong relation ("+1" instead of "2x", or vice versa) and the same percent evidently used the wrong sign (X1 instead of +1, or +2 instead of 2X).

Sixty-one percent of the students had all dots correctly labelled.

CSMP Test B, Subtest 1: Arrow Diagrams

Frequency Distribution: All Students Combined

Percent of students with indicated score



Score on Subtest 1, out of 17 items

Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 1	Standard Deviation	Percent of students with score 0-9	Percent of students with score 16-18
Q1	8.6	4.8	48	2
Q2	9.0	5.0	52	9
Q3	11.6	4.4	33	20
Q4	14.3	2.9	7	42
Q5	15.4	2.2	3	64
All Students Combined	12.3	4.6	25	43

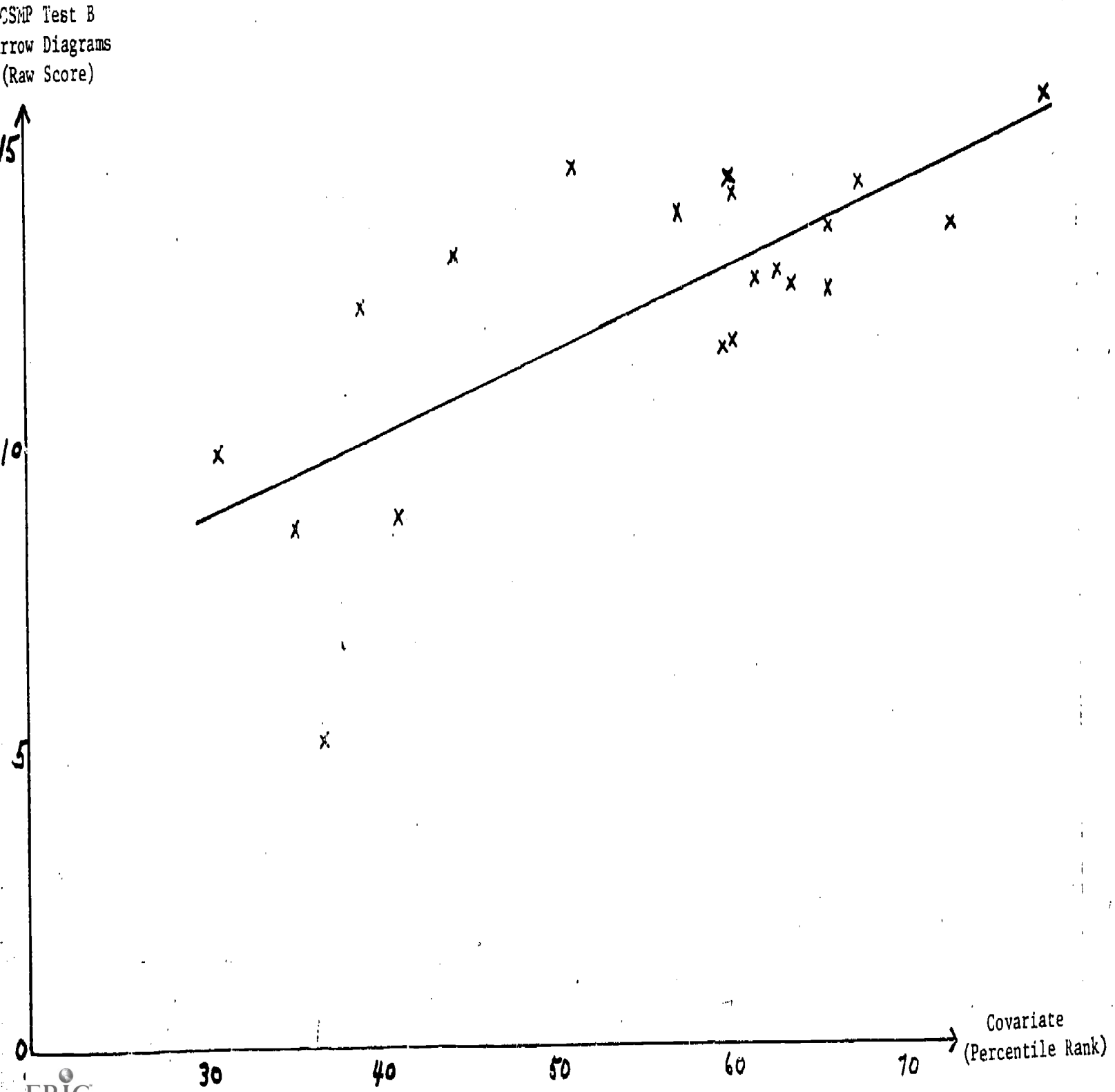
[1]Q1 means the lowest quintile, i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

Correlations Between Subtest 1 and Other Tests

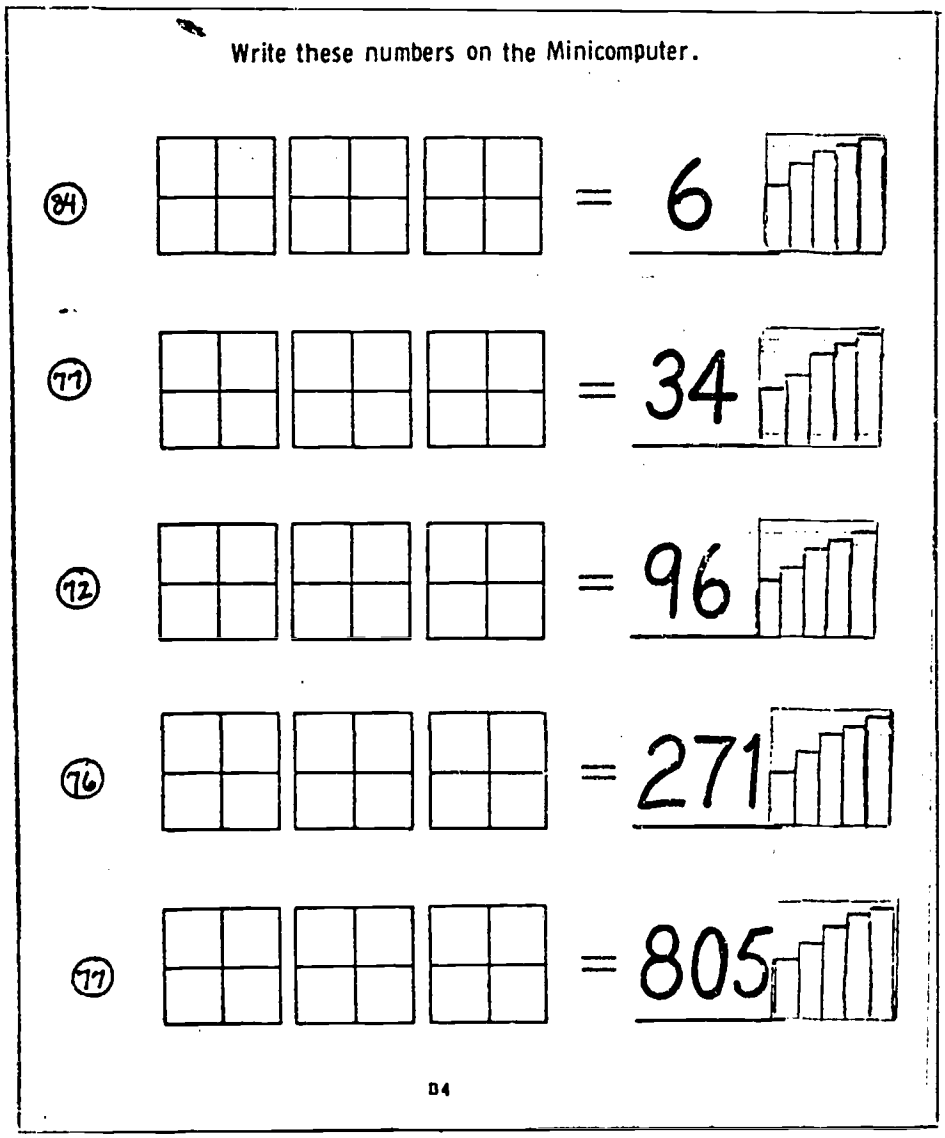
Correlations with (number of students):		
	Kuhlmann-Anderson:	.72 (153)
	SAT Reading:	.58 (97)
	CPT Reading:	.24 (98)
	Gates McGinitie:	.52 (62)
	MANS, Test A:	.71 (143)
	MANS, Test B:	.69 (313)
	CTBS Math:	.78 (241)

Reliability/Homogeneity (KR20)=.91(671)

CSMP Test B: Arrow Diagrams Versus Covariate
Second Grade Class Means



CSMP Test B, Subtest 2, Minicomputer

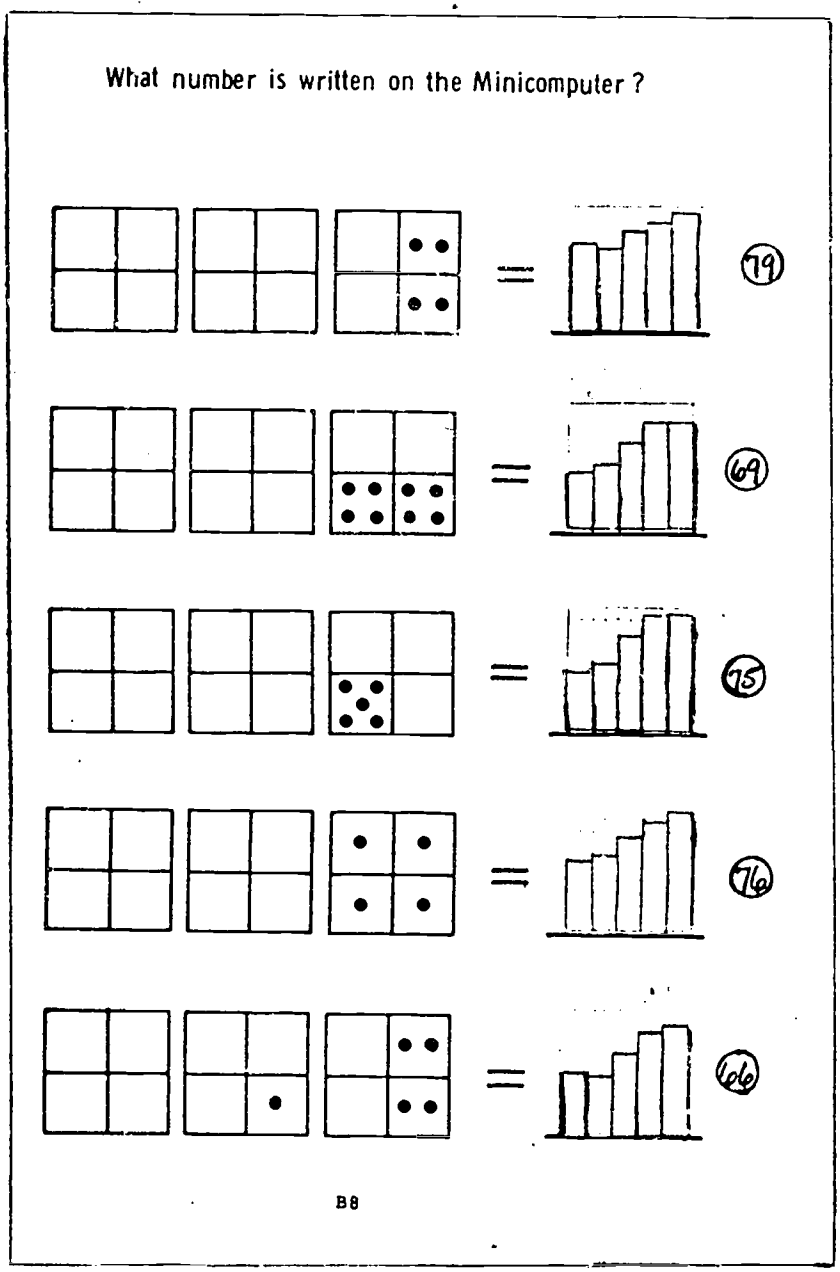


Between 5 and 13 percent of the correct answers to the above items were in what might be called "non-standard" form (for example "6" can be shown as a 4 and a 2, or as six 1's, or as a 4 and two 1's, etc.)

An average of 3 percent of the responses were wrong because of an error on one board and another 3 percent involved a digit reversal of some kind.

Ninety-six percent of the students tried all items; 1 percent omitted all items.

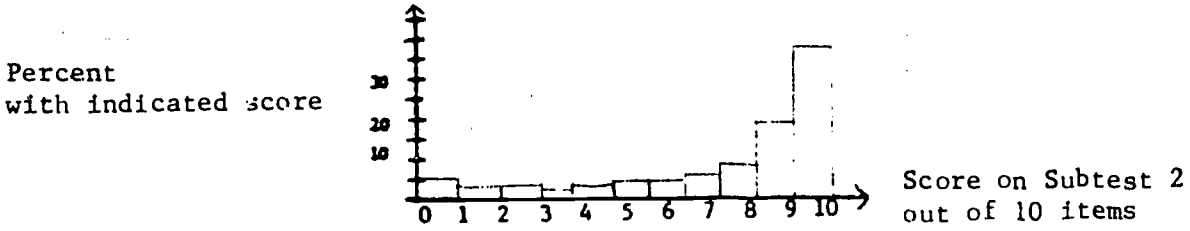
CSMP Test B, Subtest 2, Minicomputer (cont'd)



Ninety-four percent of the students attempted all items; about 4 percent omitted all. Incorrect answers were not readily categorizable except that the most common incorrect responses for the second question were "8" and "84" (instead of "12") and for the third question "5" instead of "10").

CSMP Test B, Subtest 2: Minicomputer

Frequency Distribution: All Students Combined



Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 2	Standard Deviation	Percent of students with score 0-6	Percent of students with score 9,10
Q1	5.2	3.7	54	25
Q2	6.0	3.6	52	37
Q3	7.7	2.8	25	57
Q4	8.7	2.0	14	76
Q5	9.3	1.6	5	88
All Students Combined	7.5	3.1	28	58

[1]Q1 means the lowest quintile, i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

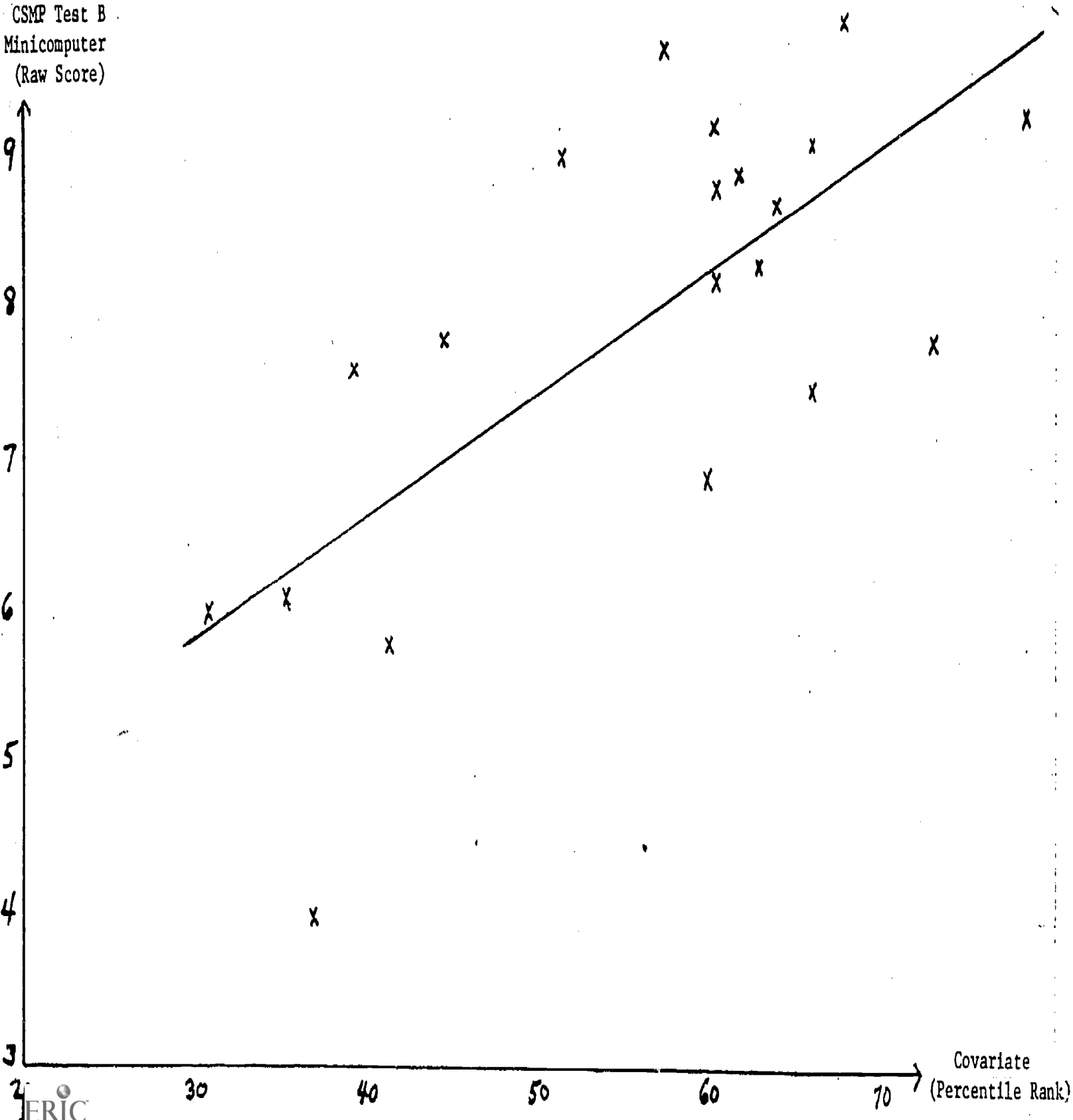
Correlations Between Subtest 2 and Other Tests

Correlations with (number of students):

Kuhlmann-Anderson:	.62 (153)
SAT Reading:	.53 (97)
CPT Reading:	.17 (98)
Gates McGinitie:	.38 (62)
MANS, Test A:	.64 (143)
MANS, Test B:	.58 (313)
CTBS Math:	.65 (241)

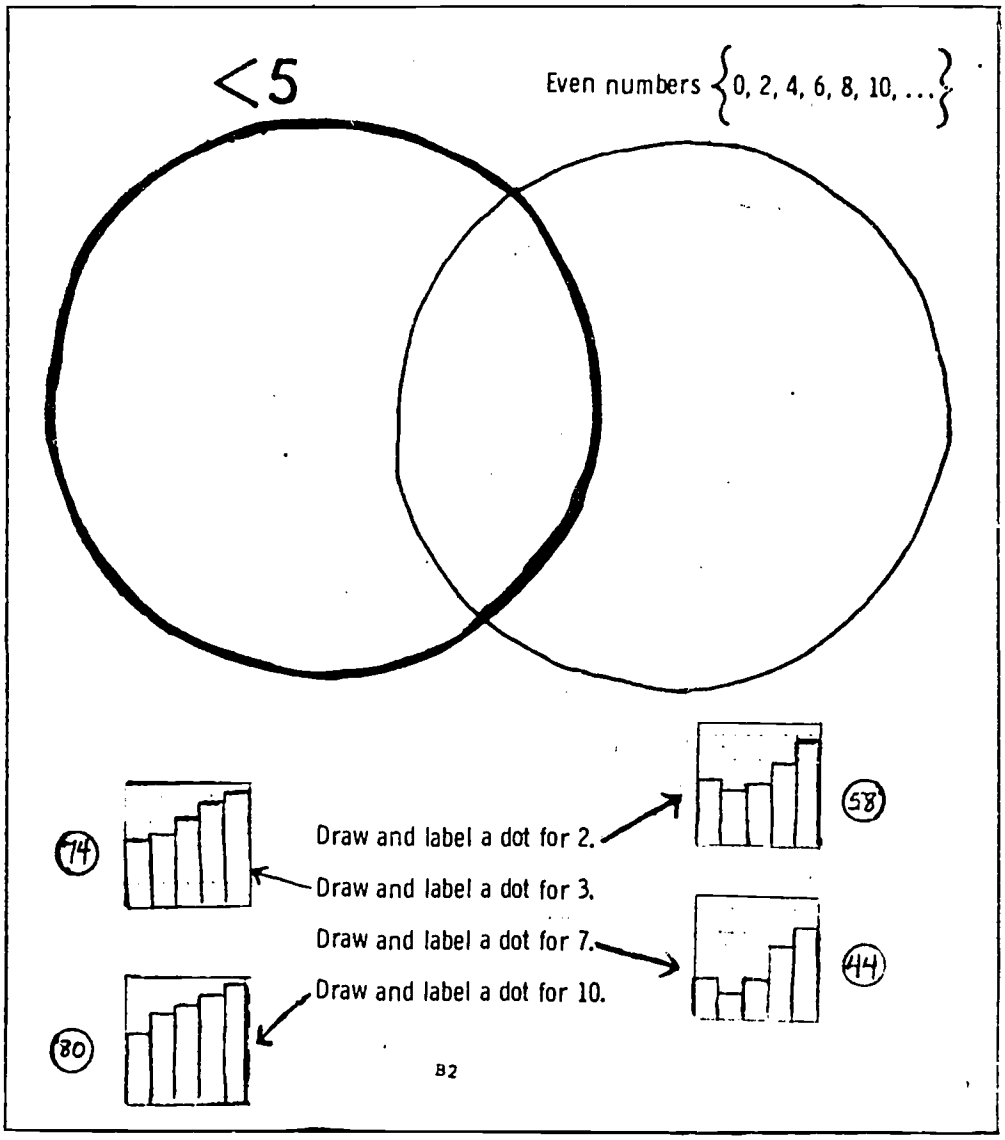
Reliability/Homogeneity (KR20)=.90(671)

CSMP Test B: Minicomputer Versus Covariate
Second Grade Class Means



CSMP Test B, Subtest 3, String Pictures (cont'd)

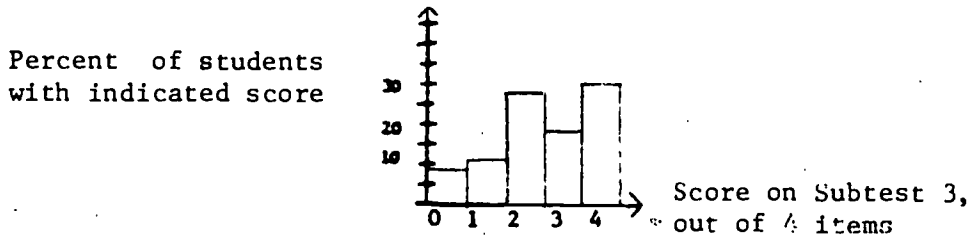
"The red string is for numbers which are less than 5 and the blue string is for even numbers like 0, 2, 4, 6, 8, 10 and so on. At the bottom it says, 'Draw and label a dot for 2.' You have to find where the dot for 2 goes and then write '2' beside it. Then do the same for the other numbers."



Averaging across the four items, 9% had the given number misclassified with regard to the blue string dimension only (e.g. the dot for 2 was placed correctly within the red string but incorrectly outside the blue string), 14% had the given number misclassified with regard to the red string dimension only, and less than 5% of the students had the given number misclassified on both categories.

CSMP Test B, Subtest 3: String Pictures

Frequency Distribution: All Students Combined



Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 3	Standard Deviation	Percent of students with score 0-2	Percent of students with score 4
Q1	2.0	1.3	65	18
Q2	2.0	1.3	66	14
Q3	2.2	1.4	56	21
Q4	3.0	1.2	36	48
Q5	3.5	0.9	17	69
All Students Combined	2.6	1.3	49	32

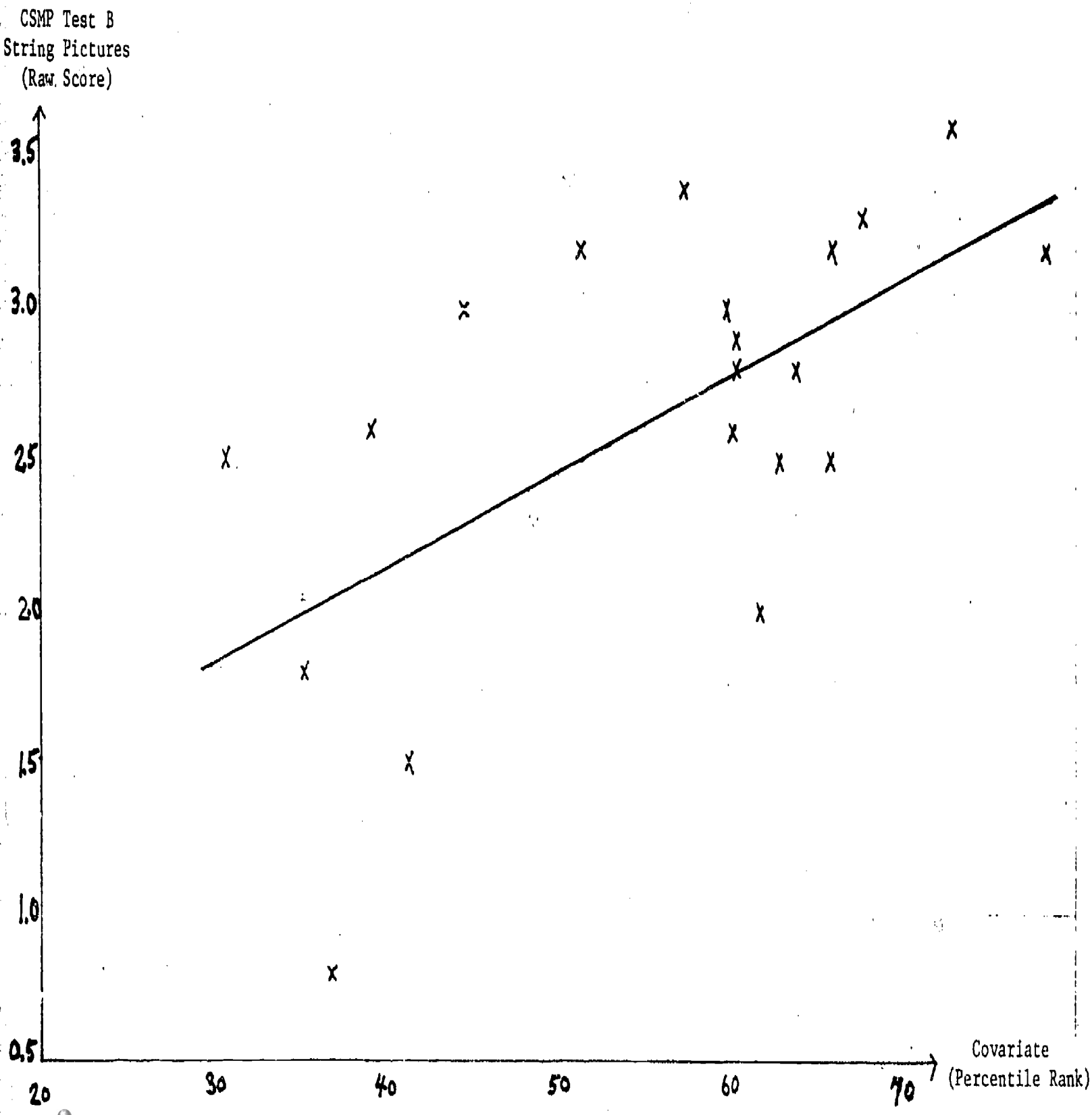
[1]Q1 means the lowest quintile, i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

Correlations Between Subtest 3 and Other Tests

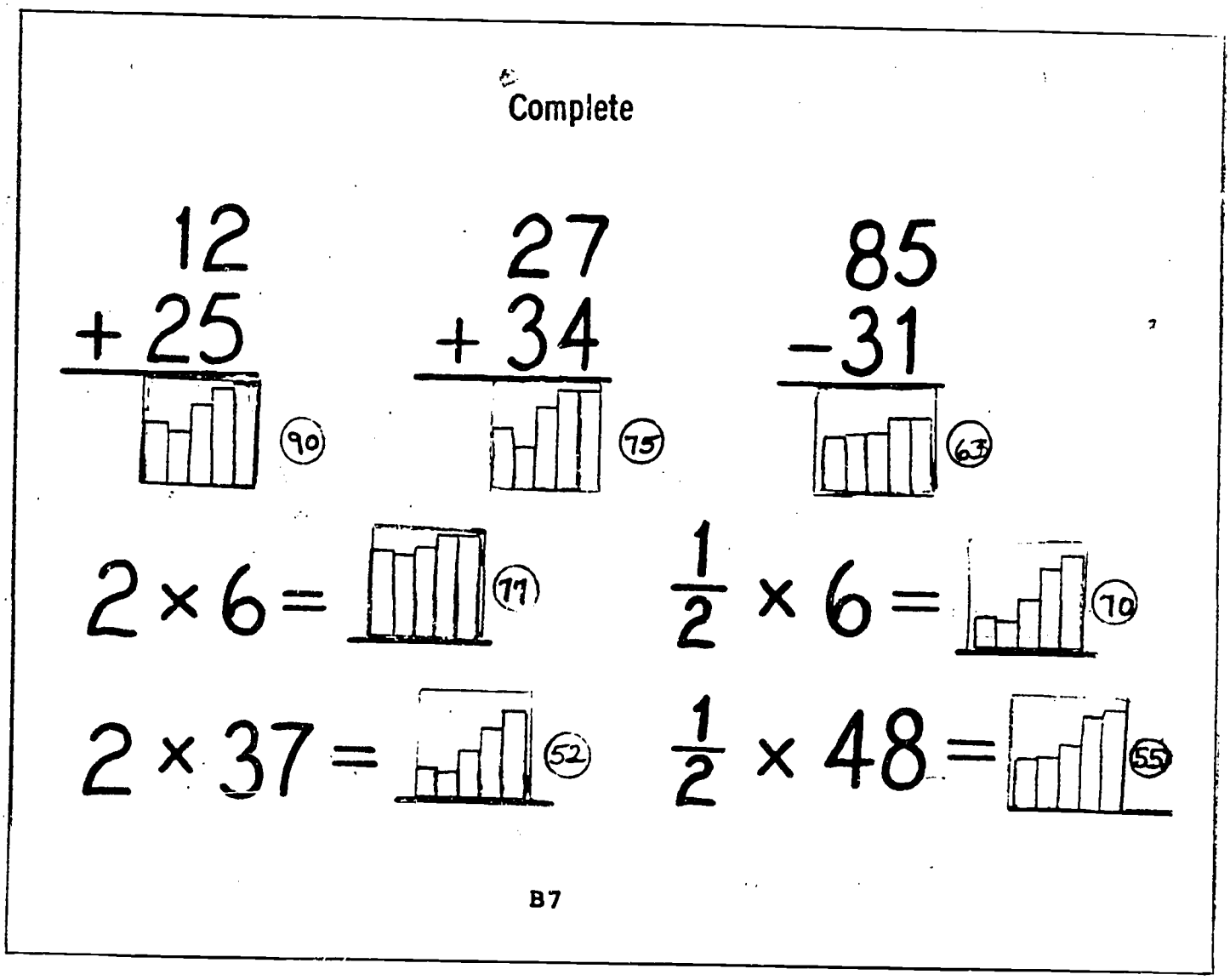
Correlations with (number of students):	Kuhlmann-Anderson:	.51 (153)
	SAT Reading:	.54 (97)
	CPT Reading:	.10 (98)
	Gates McGinitie:	.56 (62)
	MANS, Test A:	.67 (143)
	MANS, Test B:	.50 (313)
	CTBS Math:	.50 (241)

Reliability/Homogeneity (KR20)=.63(671)

CSMP Test B: String Pictures Versus Covariate
Second Grade Class Means



CSMP Test B, Subtest 4, Computation

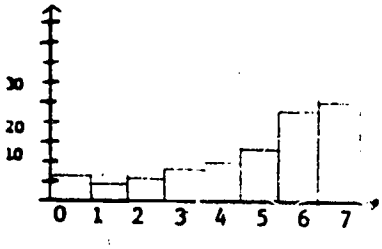


There were three incorrect responses given by at least 5 percent of the students. Two of them had to do with misreading the operation signs; adding (85+31) instead of subtracting and adding (2+6) instead of multiplying. The other had to do with carrying, in the second addition problem, where many responses of "51" or "511" were given.

CSMP Test B, Subtest 4: Computation

Frequency Distribution: All Students Combined

Percent of students with indicated score



Score on Subtest 4, out of 7 items

Summary Statistics by Ability Group

Ability Group[1]	Mean: Subtest 4	Standard Deviation	Percent of students with score 0-3	Percent of students with score 6, 7
Q1	3.4	2.3	49	20
Q2	3.1	2.1	56	14
Q3	4.3	2.2	32	42
Q4	5.7	1.4	8	63
Q5	6.1	1.2	2	79
All Students Combined	4.8	2.1	26	50

[1]Q1 means the lowest quintile, i.e. this is the set of students whose percentile rank, on whatever ability test was used as a covariate, was less than 20. Q2 is the set of students whose percentile rank was from 20 to 39, etc., and Q5, the highest quintile is for students whose percentile rank was 80 or more.

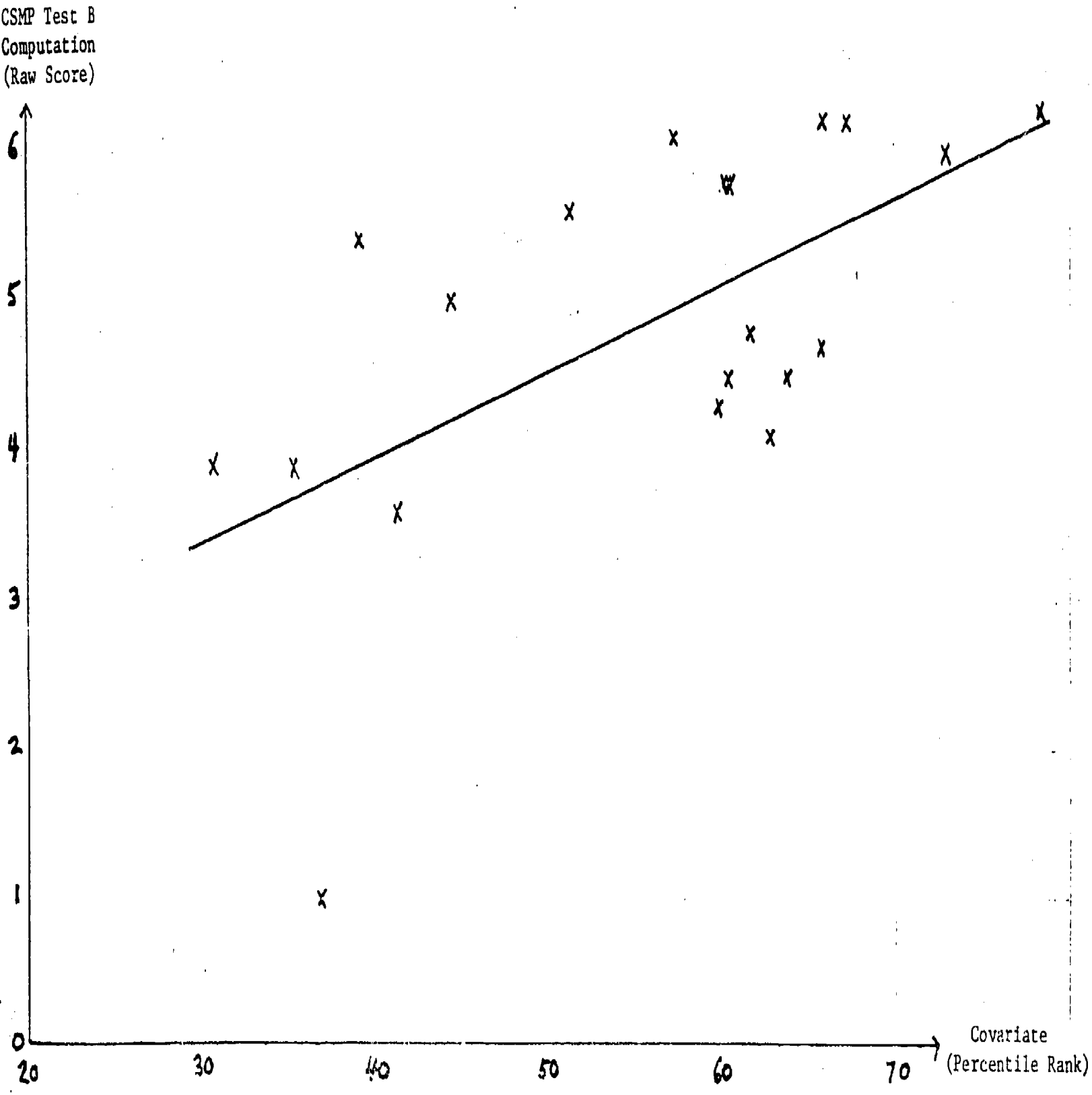
Correlations Between Subtest 4 and Other Tests

Correlations with (number of students):

Kuhlmann-Anderson:	.69 (153)
SAT Reading:	.58 (97)
CPT Reading:	.32 (98)
Gates McGinitie:	.47 (62)
MANS, Test A:	.77 (143)
MANS, Test B:	.70 (313)
CTBS Math:	.77 (241)

Reliability/Homogeneity (KR20)=.80(671)

CSMP Test B: Computation Versus Covariate
Second Grade Class Means



CSMP-Specific Test B, Intercorrelations (n=699)

	Subtest 1	Subtest 2	Subtest 3
Subtest 1: Arrow Diagrams			
Subtest 2: Minicomputer	.72		
Subtest 3: String Pictures	.49	.44	
Subtest 4: Computation	.77	.71	.51

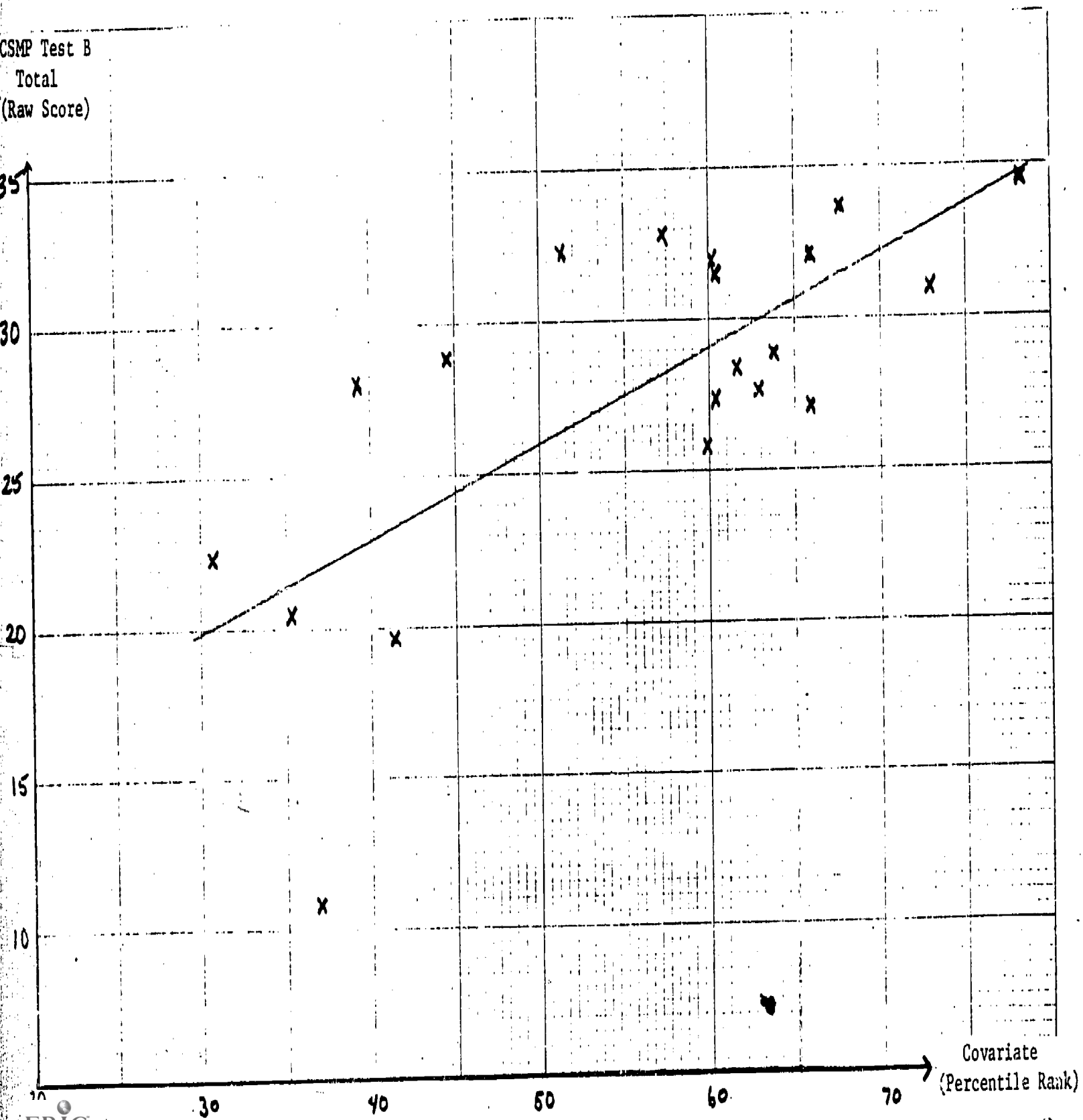
CSMP-Specific Tests A and B, Correlations (n=111)

		Test B				
		Arrow Diagrams	Minicomputer	String Pictures	Computation	Total
Test A	Arrow Diagrams	.62	.52	.38	.66	.69
	Minicomputer	.52	.56	.41	.58	.63
	Integers	.30	.28	.27	.35	.36
	Total	.63	.59	.46	.69	.73

Correlations Between Total Score Test B and Other Tests

Correlations with (number of students):	Kuhlmann-Anderson:	.74 (153)
	SAT Reading:	.64 (97)
	CPT Reading:	.28 (98)
	Gates McGinitie:	.53 (62)
	MANS, Test A:	.79 (143)
	MANS, Test B:	.74 (313)
	CTBS Math:	.82 (241)

CSMP Test B: Total Score Versus Covariate
Second Grade Class Means



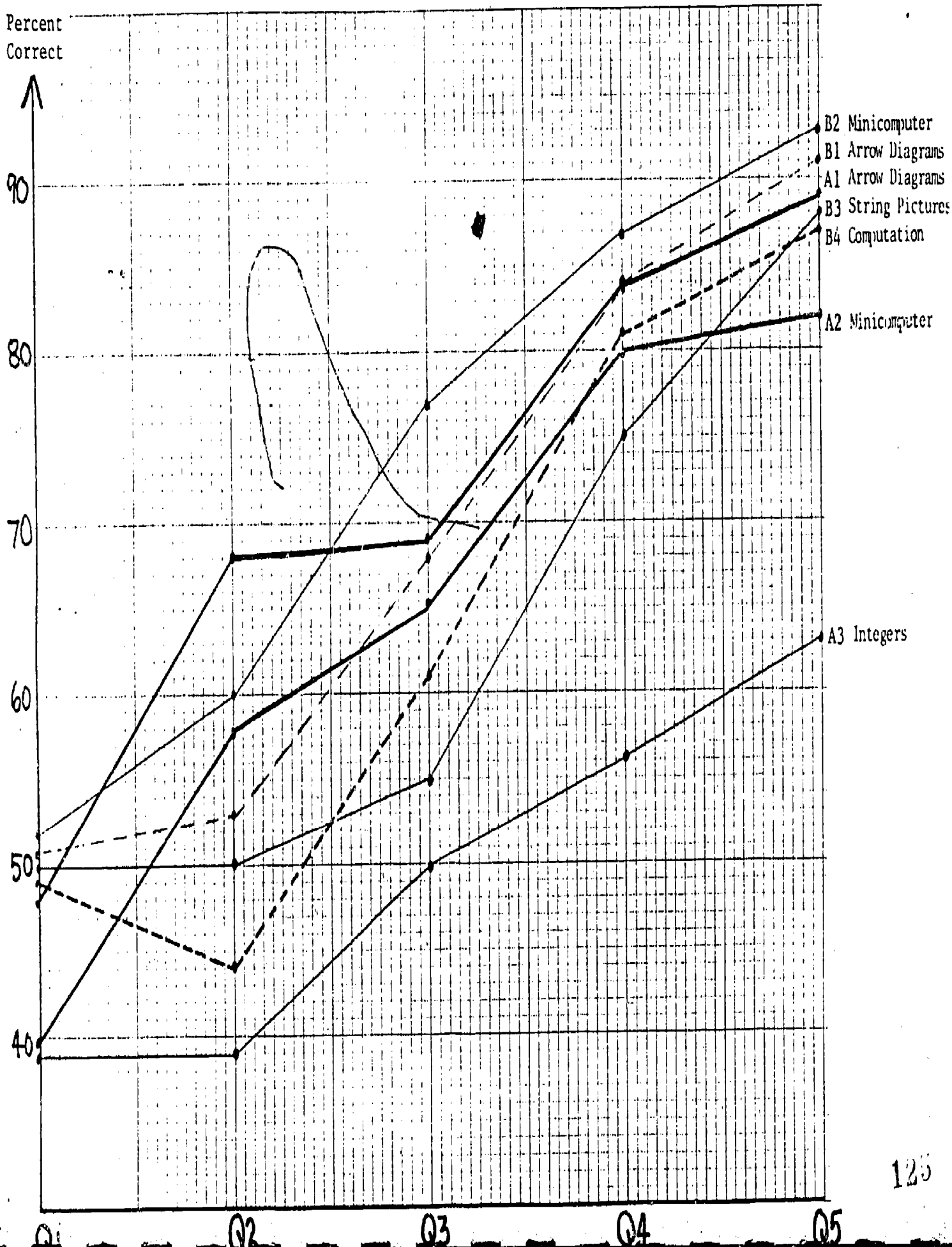
Commentary

The figure on the next page shows a graph for each of the seven subtests of CSMP-Specific Tests A and B. For each graph, there are dots which show the mean percent correct on that subtest for the set of students in Q1 (with percentile rank on covariate <20), in Q2, in Q3, in Q4 and in Q5. With the exception of subtest B3, the line segments joining the Q4 dots to the Q5 dots have a fairly gentle slope - they do not go much higher. However, for the line segments joining the Q3 dots to the Q4 dots, there is an almost uniformly large increase in the percent correct. This means that there is very little difference in performance between the two highest ability groups, but between those two groups and Q3, the middle ability group, there is a dramatic difference, and one is struck by the regularity of the graphs in this regard. It may be that somewhere around the 60th percentile there is some type of break in the distribution; above this point students do very well, usually over 80% correct, while the middle group of students are getting between 50 and 70% correct (always excepting subtest A3, Integers).

Below Q3, the graphs lose their regularity. The Q2 dots are always below the Q3 dots of course, but there is much variation in the size of the difference. Between students in the lowest quintile, Q1, and students in the next-to-lowest quintile, Q2, there are four subtests on which there is virtually no improvement in performance; it is not until Q3 that one sees an improvement in performance.

Mean Percent Correct by Quintile,
CSMP-Specific Subtests

95/D



Appendix E

CSMP-Specific Individually Administered Test, Second Grade

This test was administered on a one-to-one basis to students from 37 CSMP second grade classes in 6 sites. Four students, of varying but representative ability levels, were selected from each class. The test was administered by special testers, trained by the CSMP coordinator at each site, and required about 15 to 20 minutes per student.

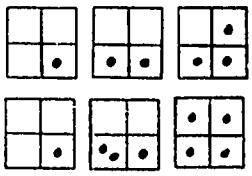
The format for administering and recording responses from the test was quite lengthy and is not reproduced in full here. Instead a rather shortened version of the test items and materials is given, together with the percent of the students who got each item correct. The testers were encouraged to explain the questions as fully as possible, with alternative wording where appropriate and, with the first two or three of a sequence of similar tasks, to show the student how to do the question if necessary.

The percentage correct is based on all 148 students who took the test. About half these students were from classes from which covariate scores (from reading or general ability tests) were available. For these classes it was possible to determine the degree to which the selected students were representative of the classes from which they were chosen. The mean percentile rank of the selected students was between 5 and 10 points higher than the mean percentile rank for all students in these classes. Hence these students tend to be slightly higher in ability and, to the extent to which one might make comparisons between data from these individually administered tests and the group administered CSMP Tests A and B, one should bear this in mind.

On the last page of this Appendix, after the test items and item data have been presented, some summary data for each of the subtests (Minicomputer, String Pictures, and Arrow Diagrams) is presented.

Subtest 1: Minicomputer

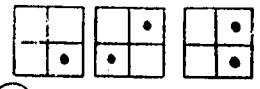
Note: after each question in sequence below, if the student was unable to do the question (i.e. arrive at the configuration shown at the right), the tester showed the student how to do it. Hence each new question started with the same configuration for each student.



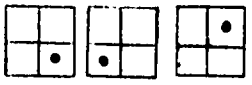
1. "Show me one hundred and thirty-seven on the Minicomputer." (88)

2. "Now you add 28 to the number on the Minicomputer." (88)

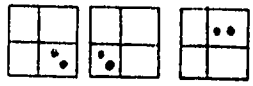
3. "Figure out what the Minicomputer says. You might have to make some plays."
   a) Made 8+2→10 play (64)
   b) Made 10+10→20 play (80)
   c) Made 20+20→40 play (79)
   d) Correctly read off the number displayed on the Minicomputer (77)



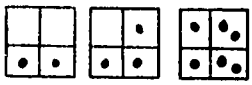
4. "Subtract 41 from that number." (84)



5. "Let's make the number twice as big. How would we multiply it by two?" (69)



Note: before each of items 6-11 below was asked, the Minicomputer was returned to the configuration shown at the right.



"I'm going to move one of the checkers to a new square. When I've done that, you tell me whether I've made the number on the Minicomputer larger, smaller, or whether it's still the same."

6a) "I'm going to take this checker (pick up a checker from the 1's square) and put it here (put it on the 4's square). Did I make the number larger, or smaller, or is it still the same?" (91)

b) (If correct) "How much bigger did I make the number?" (37)

7a) Same as 6a), but move a checker from the 8's square to the 2's square. (79)

b) (If correct) "How much smaller did I make the number?" (36)

8. Same as 6a), but take two checkers from the 4's square, put one on the 8's square and discard the other one. (46)

9. Same as 6a), but move a checker from the 200's square to the 20's square. (78)

10. Same as 6a), but move a checker from the 40's square to the 400's square. (86)

11. Same as 6a, but take a checker from the 20's square and put it, plus a new checker, onto the 10's square. (49)

12. "Now you move a checker to make the number on the Minicomputer larger. Make it larger by 1. Make it show a number that's one more than the number that it shows now." (44)

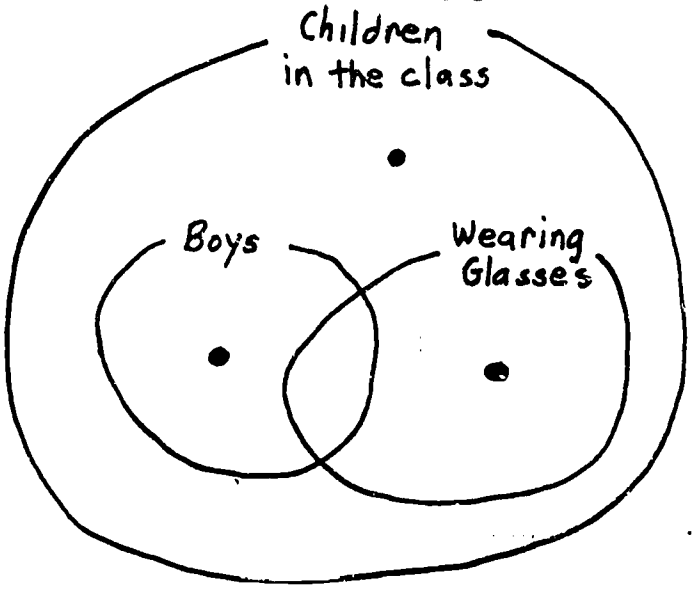
13. "Move a checker to make the number larger by 40." (25)

14. "Now make it larger by 7." (21)

15. "Now make it smaller by 3. (26)

Subtest 2: String Pictures

"Look at this string picture. (Reduced version below was originally in color.) This big string is for children in the class. The red string is for boys. And the blue string is for children who are wearing glasses."



16. "Now what I want you to do is pretend you are in this class. Put a dot for where you would go." (91)

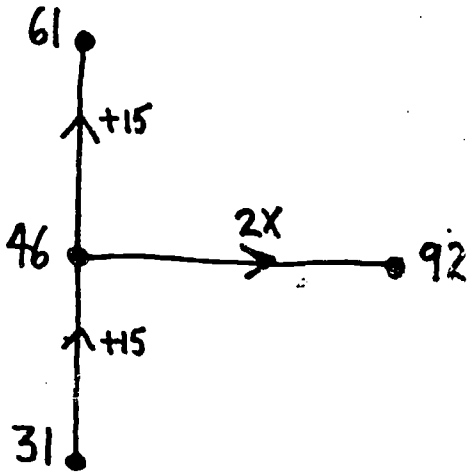
17. "What can you tell me about this dot?" (If student being tested is a boy - point to top dot; if a girl - point to left-most dot.)
    a) (If necessary) "Is it for a boy or a girl?" (85)
    b) (If necessary) "Does this child wear glasses?" (86)

18. "What can you tell me about this dot?" (Point to right-most dot.)
    a) (If necessary) "Is it for a boy or a girl?" (74)
    b) (If necessary) "Does this child wear glasses?" (93)

19. "Put a dot for me (for the tester). Where would I go?" (44)

Subtest 3: Arrow Diagrams

"Now let's look at this arrow diagram. (Reduced version, below, was oringinally in color.) If someone weren't very good at calculations they could use this diagram to help them. Let's pretend we aren't very good at figuring out number problems and see which facts we could know just by looking at the diagram."



"I'm going to give you some problems and you decide whether or not you can figure out the answer from the picture."

20a) "Could we use this diagram to figure out what 31+15 is?" (91)
b) (If "yes") "What is the answer to 31+15?" (71)

Note: for the first two or three items of this subtest the tester had the student show or showed to the student if necessary, how one would use the diagram (which arrows from which numbers, etc.).

21. "Could we use the diagram to figure out 2x92?" (66)

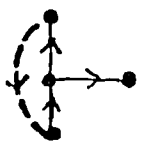
22a) "Could we use the diagram for 61-15?" (51)
b) (If "yes") "What is the answer?" (31)

23. "Could we use the diagram to figure out 31-15?" (74)

24a) "Could we use the diagram to figure out ½ of 92?" (53)
b) (If "yes") "And what is the answer?" (31)

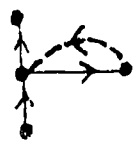
25a) "Could we use the diagram to figure out 31+30?" (54)
b) "(If "yes") "What is the answer?" (41)

26. (If correct answer given to #25) "Suppose I draw an arrow like this: What could this arrow be for?" (33)



27. (If incorrect answer given to #25, same as #26, except reverse the direction (arrow head) of the new arrow.) (24)

28. (If incorrect answer given to #24) "Suppose I draw an arrow like this: What could this arrow be for?" (35)



CSMP Individually Administered Test, Summary Statistics

	Number of Items	Mean	Standard Deviation	Correlation with:	
				Subtest 1	Subtest 2
Subtest 1: Minicomputer	20	12.4	4.7		
Subtest 2: String Pictures	6	4.7	1.1	.40	
Subtest 3: Arrow Diagrams	12	6.0	2.6	.12	.08

Correlations Between Individually Administered Tests and Other Tests

	CSMP Individually Administered Tests			
Other Tests	Minicomputer	String Pictures	Arrow Diagrams	Total
MANS A	70	.47	.31	.73
MANS B	49	.15	.31	.50
CSMP Test A: Arrow Diagrams	.50	.30	.11	.52
Minicomputer	.58	.27	.04	.56
Integers	.47	.23	.28	.53
CSMP Test B: Arrow Diagrams	.62	.28	.55	.69
Minicomputer	.66	.23	.63	.73
String Pictures	.58	.29	.41	.63
Computation	.62	.26	.50	.67

Commentary

It is not possible to directly compare students' performance on this test with their performance on the CSMP-Specific group administered tests. For the most part the questions in this individual test simply could not be asked in a paper-and-pencil setting. The questions were generally more difficult and the testers helped the students with explanations, corrections, encouragement, etc. They were also told to be lenient in the scoring protocol.

A total of 6 testers in different sites were used and each was trained by the respective local CSMP coordinator. No doubt there was some variation from site to site in the way the tests were administered and scored. Given these circumstances it was not feasible to "open up" the testing process to a deeper probing of student responses. (For example, by asking to explain his or her response, by backtracking from points of difficulty, asking progressively harder questions after success with easier ones, etc.) That would have yielded more interesting and explanatory information but would have been far beyond the scope of this study.

Rather low correlations were found between the String Pictures subtest and the group administered CSMP-Specific Tests. Some very low correlations were found between the Arrow Diagrams subtest and the subtests from CSMP-Specific Test A (though not Test B!). In particular the correlation between Arrow Diagrams (individually administered) and Minicomputer (group administered Test A) was .04. In other words there was virtually no relationship between student performance on these two tests. Whether this was a function of the variation across testers in the administration and interpretation of the test or of the different types of items - the "doing" as opposed to the "writing" - cannot be determined.

Appendix F

Item Analysis of Third Grade CSMP-Specific Tests A and B

On the following 16 pages are reproduced, in somewhat reduced form, the 8 pages which constituted Test A and the 8 pages which constituted Test B. (The actual page numbers of the test are part of that reduced page, for example A3 is page 3 of Test A.)

Three kinds of additional information are presented:

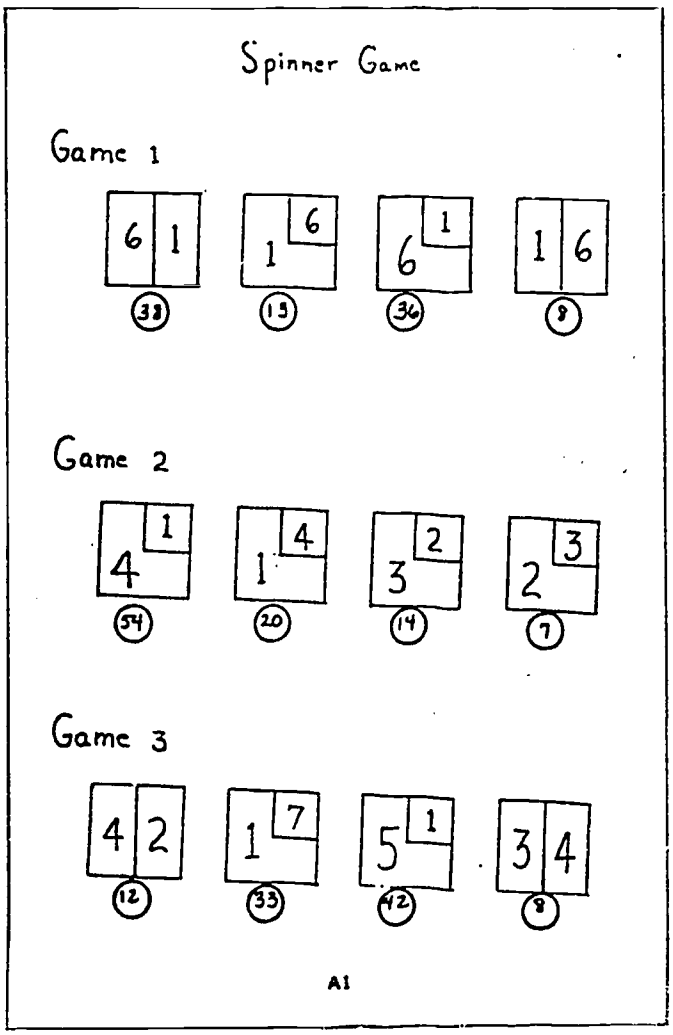
a) At the top of each page are given any directions for that page that were given by the tester to the whole class before they began that page. This was not usually considered necessary, since there were no new types of problems nor directions that were not found in various regular CSMP workbooks and worksheets. Additional explanations were given individually to students throughout the test as the need arose.

b) In the middle of the page, superimposed on the reduced test page, a circled number is given beside each test item. Except where otherwise noted, this is the percent of students who got the item correct.

c) At the bottom of the page are given any other statistics which are thought to be of interest for that page, such as: percent of students who omitted the item, correlation coefficients with scores from the Kuhlmann-Anderson test (administered previous fall), common incorrect responses, comparisons with the performance of second grade students (where applicable), etc.

The data from Test A are based on 76 students from 4 local St. Louis area classes. The mean score on the Kuhlmann Anderson test corresponds to a percentile rank of approximately 45. The data from Test B are based on 59 students from 3 classes, and their mean score corresponds to a percentile rank of approximately 56.

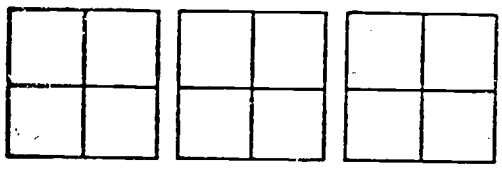
"Look at the first question at the top of the page. It says 'Game 1', and four different spinners are shown. Now listen and I'll tell you about this game. You will have a hundred spins in the game. After each spin you will get the number of points the spinner is pointing at. You think about which of the four spinners would be likely to give you the most points after a hundred spins. Then put an x on the spinner that you would choose. Then do the same thing for the other games."

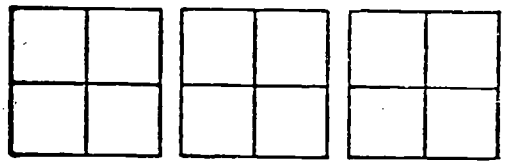


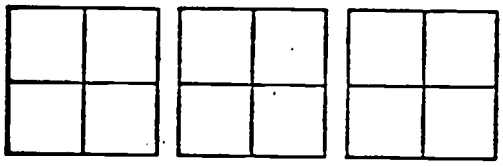
In Game 2 and Game 3, it is interesting to note that while the most popular choice was the best answer, the next most popular choice was the worst answer.

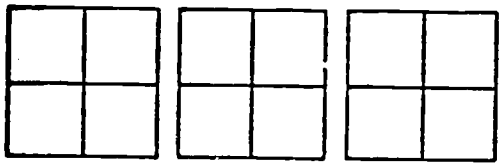
The mean correlation between the three test items and the Kuhlmann Anderson test was .31.

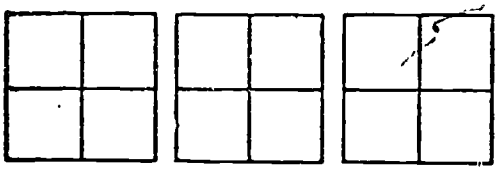
Write these numbers on the Minicomputer.

 = 6 (95)

 = 34 (85)

 = 96 (80)

 = 271 (79)

 = 805 (80)

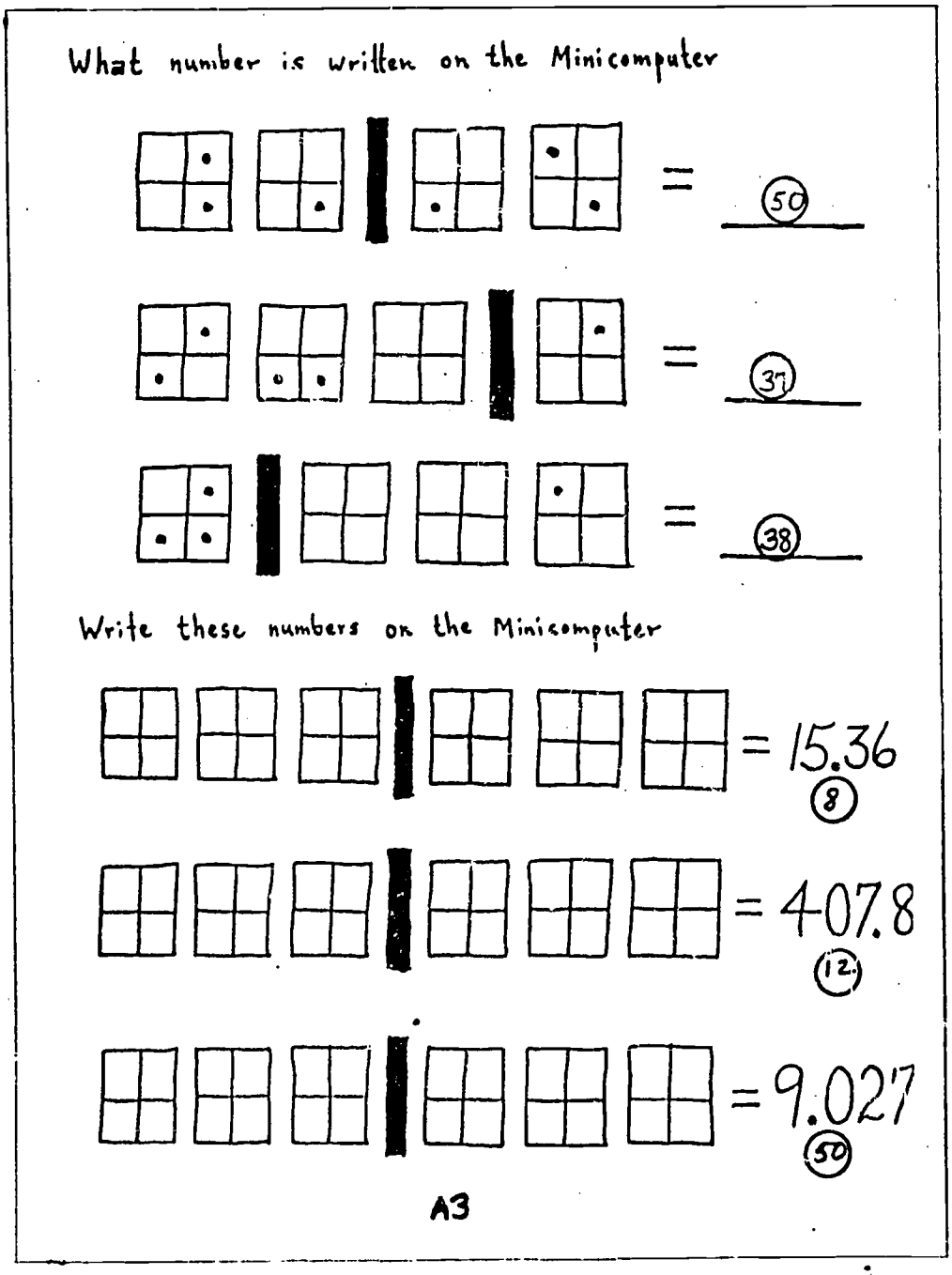
A2

The mean percent correct for these 5 items was 84. The same page was given to second grade students (CSMP Test B) and the mean percent correct for them was 78.

Approximately 12 percent of the correct answers were given in what might be called "non-standard" form. For example, 6 is usually shown on the Minicomputer with a checker on the four and two's place, but it can also be shown with 3 checkers on the two's place, and by various other configurations.

The mean correlation between the five test items and the Kuhlmann Anderson test was .50.

Third Grade, CSMP Test A (cont'd)

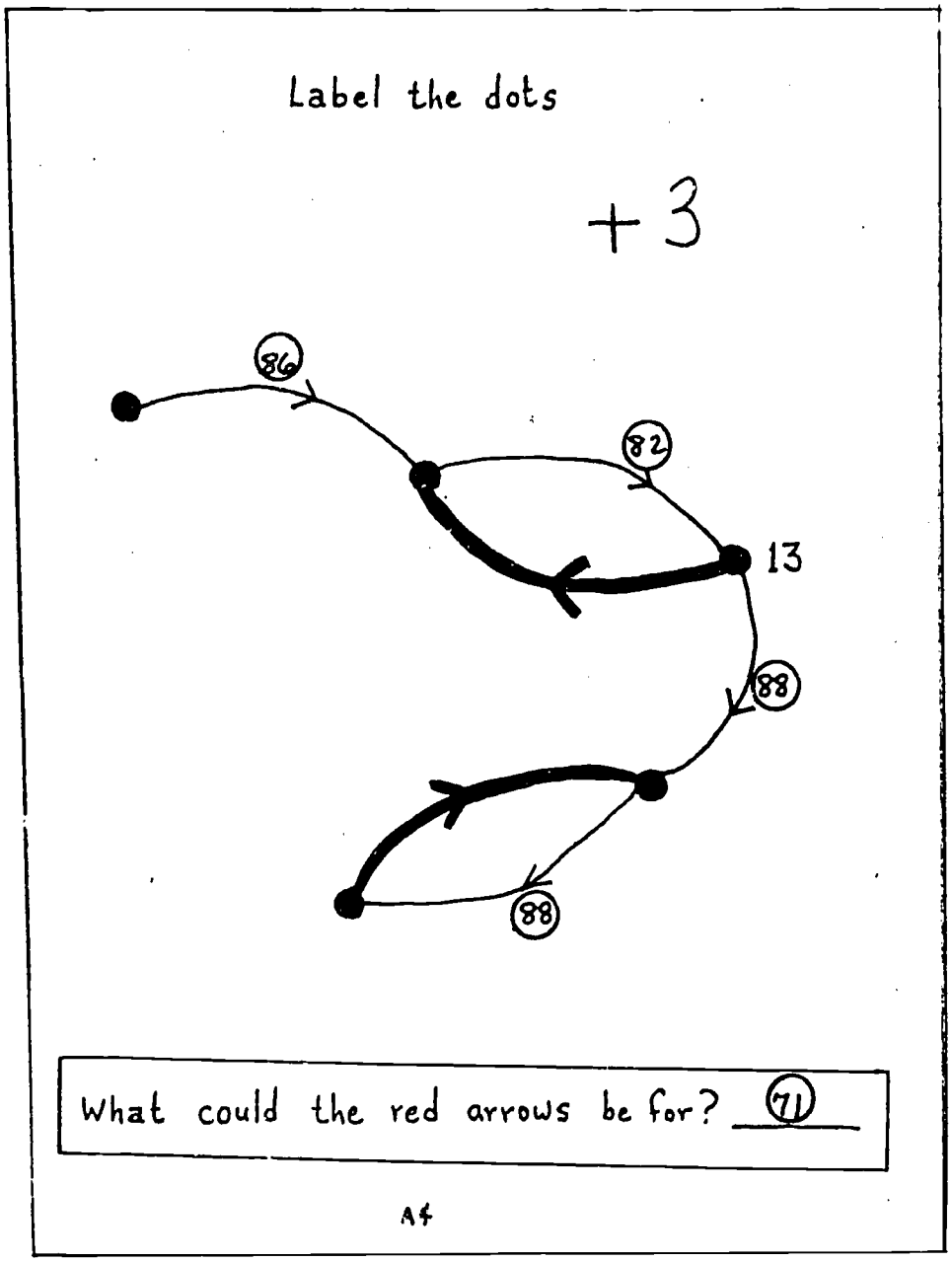


For the six items, an average of 13% of the students showed the correct sequence of digits, but placed them incorrectly in relation to the decimal point or the decimal bar (colored green on the original). This was more common in the top three items.

For the six items, an average of 26% of the students left out a digit, or reversed the digits. This usually involved the digit "0"; more than half the students had this problem with the fourth and fifth items, where they most often showed the numbers 15.036 and 407.008 respectively.

The correlation between this set of items and the Kuhlmann-Anderson Test was .34.

Third Grade, CSMP Test A (cont'd)



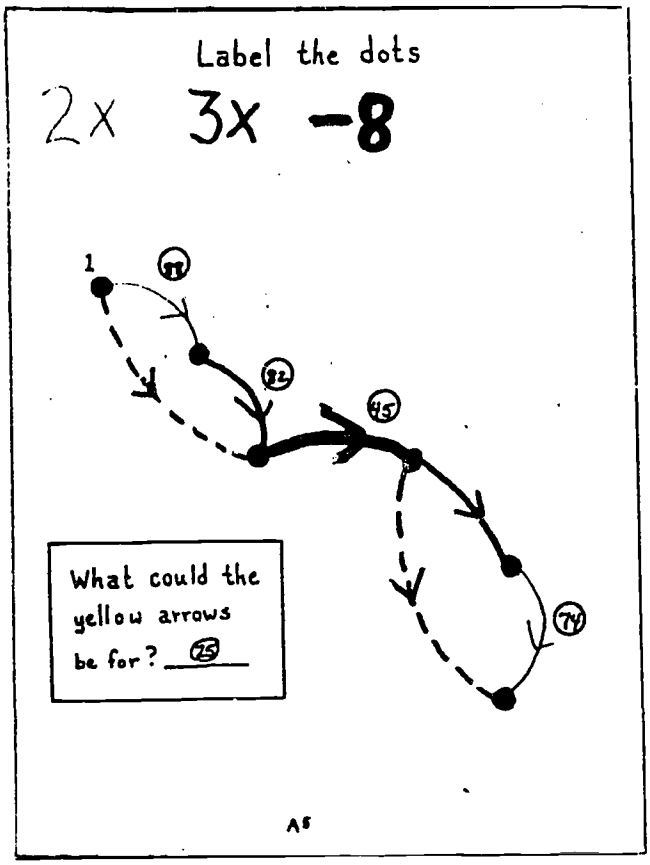
The red arrows in the original text are depicted here by the heavier arrows.

The item statistics are given for each arrow and each was scored independently. An item (arrow) was counted as correct if the dots at the end of the arrow "fit" the relation defined by the arrow.

The mean percent correct for the 5 items was 83. The same page was also given to second grade students (CSMP Test A) and the mean percent correct was 68.

The correlation between this set of items and the Kuhlmann Anderson test was .42.

Third Grade, CSMP Test A (cont'd)



Different colors were used for different valued arrows; the dotted arrow was originally yellow.

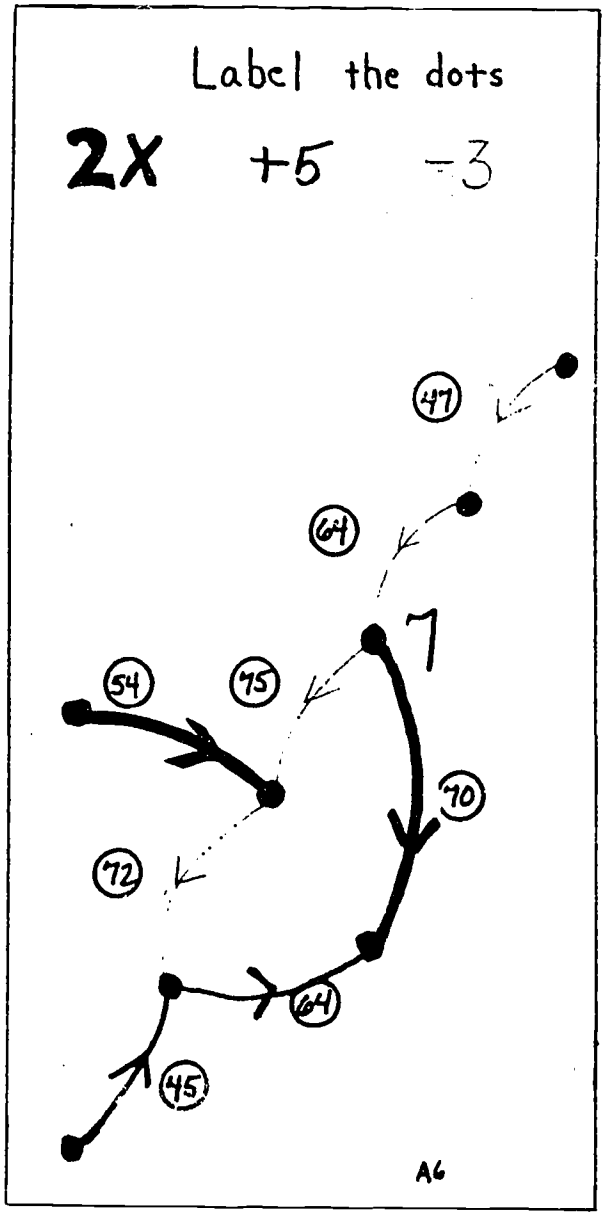
Each arrow was scored independently, as on the previous page. The percent correct for one of the red arrows has not been indicated because that item was treated in a special way. Unfortunately an error was made in producing this page for test administration and a student doing the question correctly would have dots which were labelled, successively, "1", "2", "6" and then "$\hat{2}$". Then he would have to multiply $\hat{2}$ by 3, a process not attended to in the curriculum. Therefore when analyzing responses to this particular arrow, students were divided into two groups: those who had labelled the previous dot $\hat{2}$ or some other negative number (52% of the students); and those who had labelled it as a non-negative number (48%), evidently making a mistake previously. Of the first group, 75% were correctly able to multiply their negative number by 3. Of the second group, only 52% were able to multiply their non-negative number by 3. What is surprising is not that more of the first group got it right than in the second group (they were probably better students) but that such a high percent of this first group was able to multiply a negative number.

The most common erroneous answer for the question about the yellow arrow was "5x".

Each of the last four parts to this item (including the "yellow arrow" question) were omitted by an average of 12 percent of the students.

The correlation between this set of 6 items and the Kuhlmann Anderson test was .26.

Third Grade, CSMP Test A (cont'd)



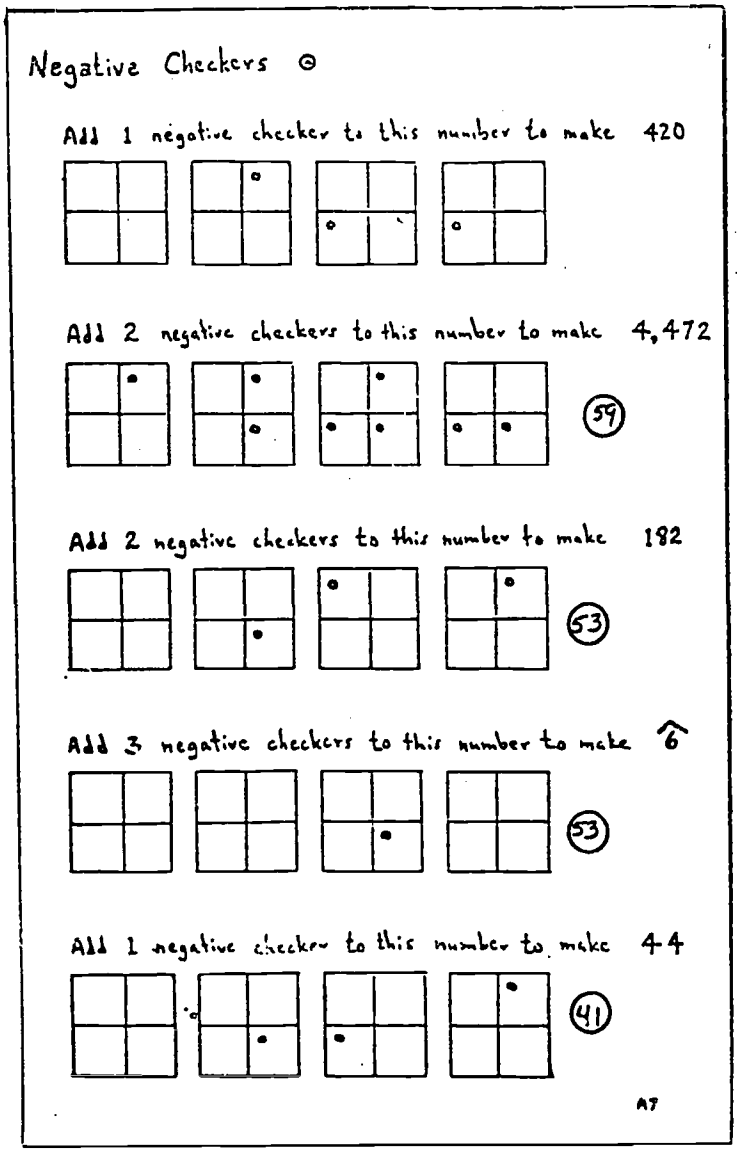
Each arrow was again scored independently. Across items (arrows), the average percent of students who omitted an item (i.e., at least one end of an arrow not labelled) was 13, with the most omits occuring with the top (+5) arrow.

For each arrow, a check was made on how often students evidently used the wrong one of the three given arrows, and on how often the student evidently reversed the arrow (for example treating the +5 arrow as -5, or the 2x arrow as $\frac{1}{2}x$). In only two instances did more than 5% of the responses have one of these errors and both were arrow reversals: in the left-most 2x arrow (16%) and in the lowest -3 arrow (25%).

The correlation between this set of items and the Kuhlmann Anderson test was .55.

Third Grade, CSMP Test A (cont'd)

The first question was done together with the class, as an example. Then "The others on this page are the same, except that you must read to see how many negative checkers to use and what number you are to make."



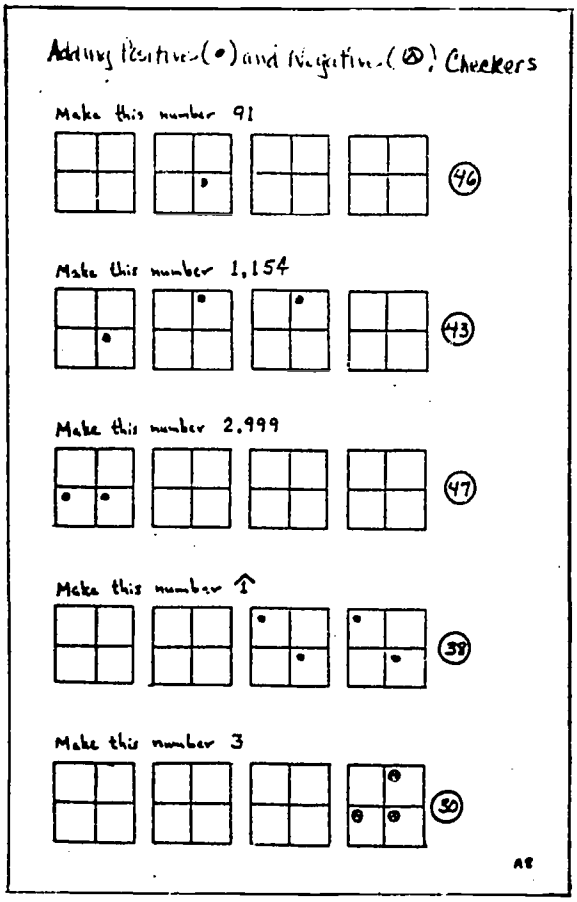
For the four items, the average percent of students who omitted an item was 16.

Most of the wrong answers were answers in which the correct number of checkers was used but the number displayed on the Minicomputer was not correct. Never more than 5 percent of the students used the wrong number of negative checkers to show the required number.

The correlation between this set of 4 items and the Kuhlmann Anderson test was .45.

Third Grade, CSMP Test A (cont'd)

"You can see that this page is like the last one except that you can use positive or negative checkers. Or you can use both in the same problem if you want."

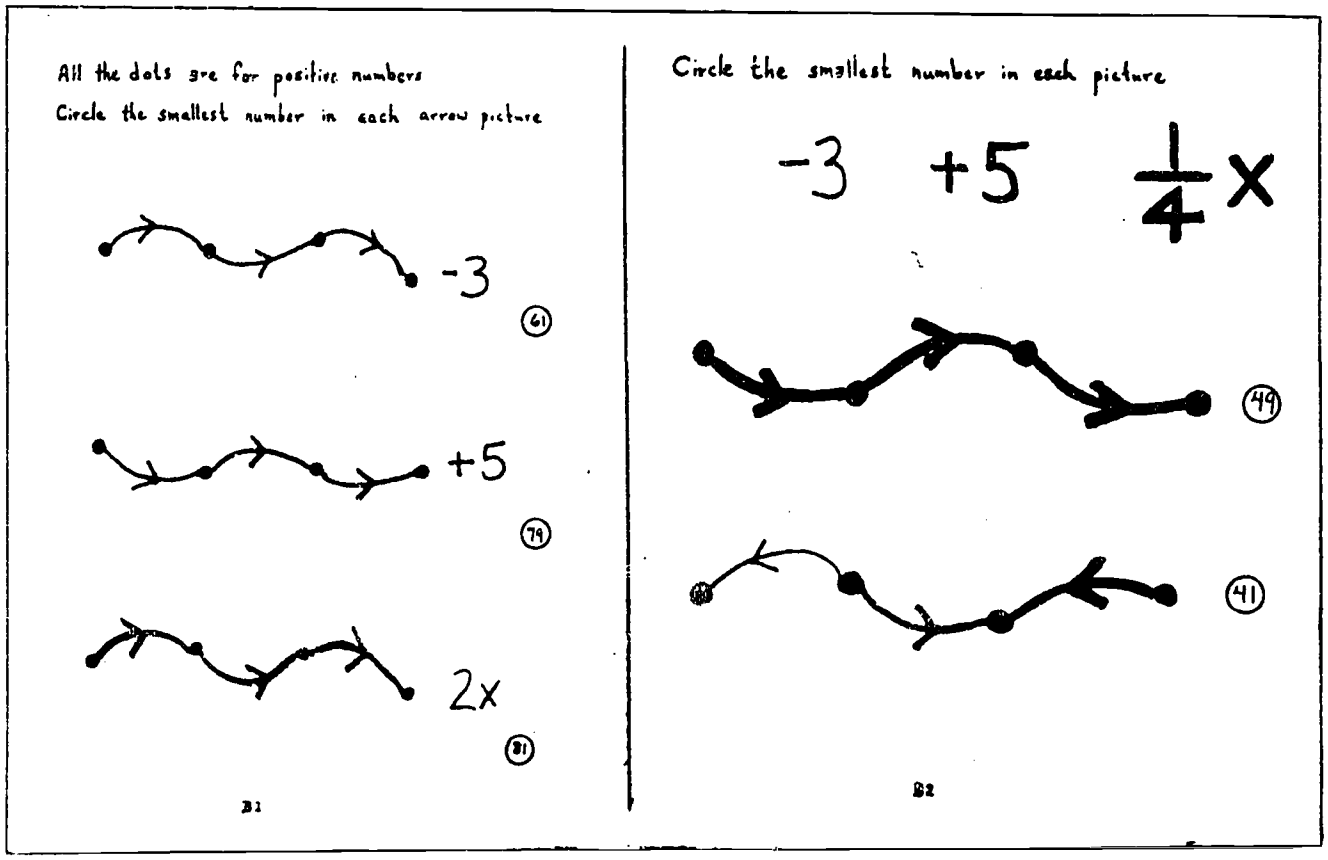


Students generally used one of two strategies for solving these problems. First, they could "cancel" the given number by pairing a negative checker with each given (positive) checker and then merely displaying the required number. (For the first item one would put a negative checker on 100 and then positive checkers on 80, 10 and 1.) Or second, they could figure out the difference between the displayed number and the required number and then add the right kind of checkers to make up this difference. (For the first item 100 is given and 91 is required - hence one needs to subtract 9 - hence one would put negative checkers on 8 and 1 or put a negative checker on 10 and a positive checker on 1.) The first strategy was the more popular, especially with the third item (which could otherwise have been solved by merely putting a negative checker on 1).

For each item, an average of 19% of the students omitted the item.

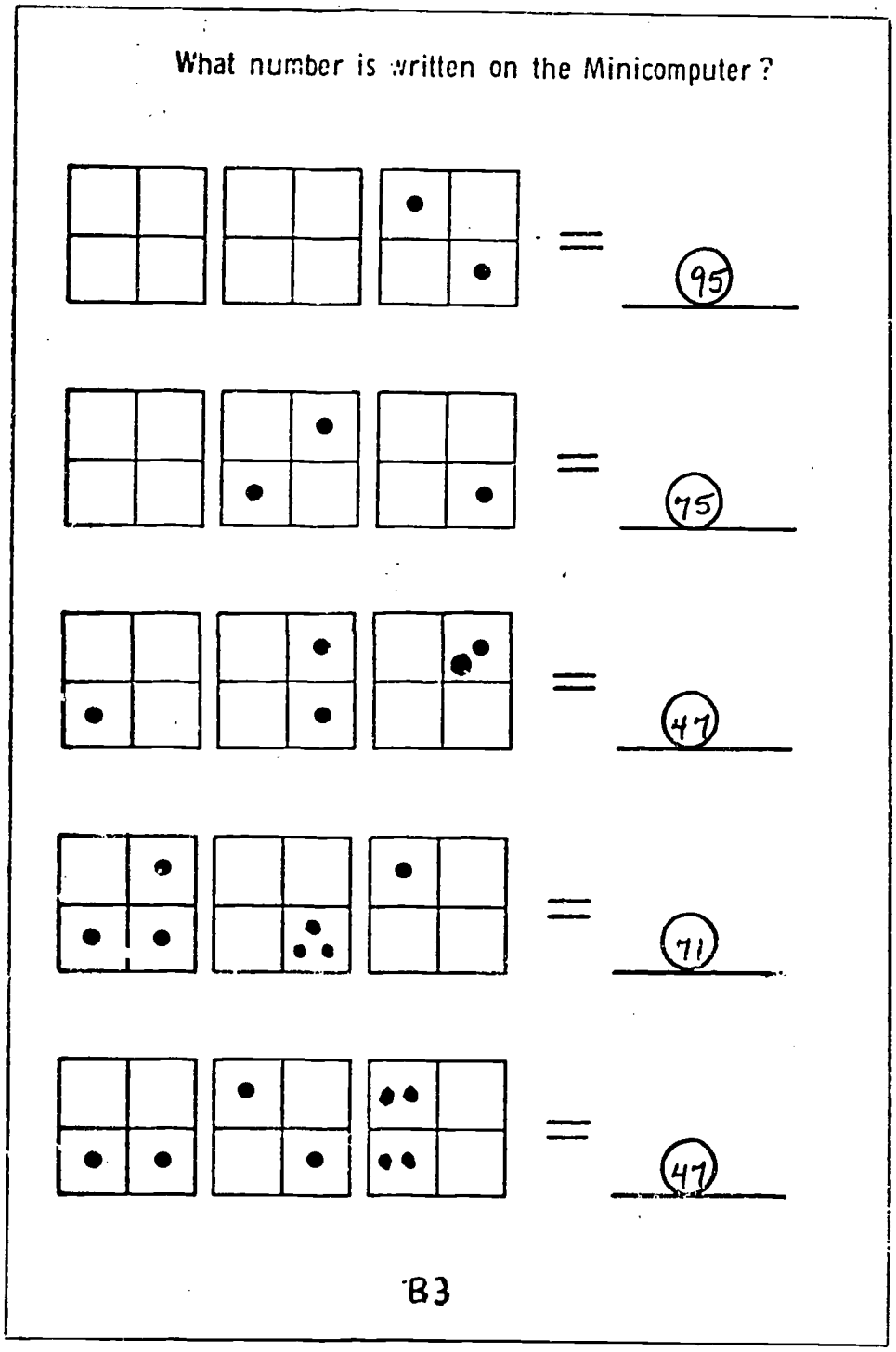
Third Grade, CSMP Test B

After reading the directions at the top of the page: "Look at the first arrow picture - the blue one. One of these four dots is for a positive number that is smaller than any of the others. You have to figure out which one it is and draw a circle around it. You don't have to label any of the dots, but you can if you want to. Then do the same for the other pictures and for the ones on the next page."



In about half the correct responses the students had labelled some of the dots to help them; in the other half no dots were labelled.

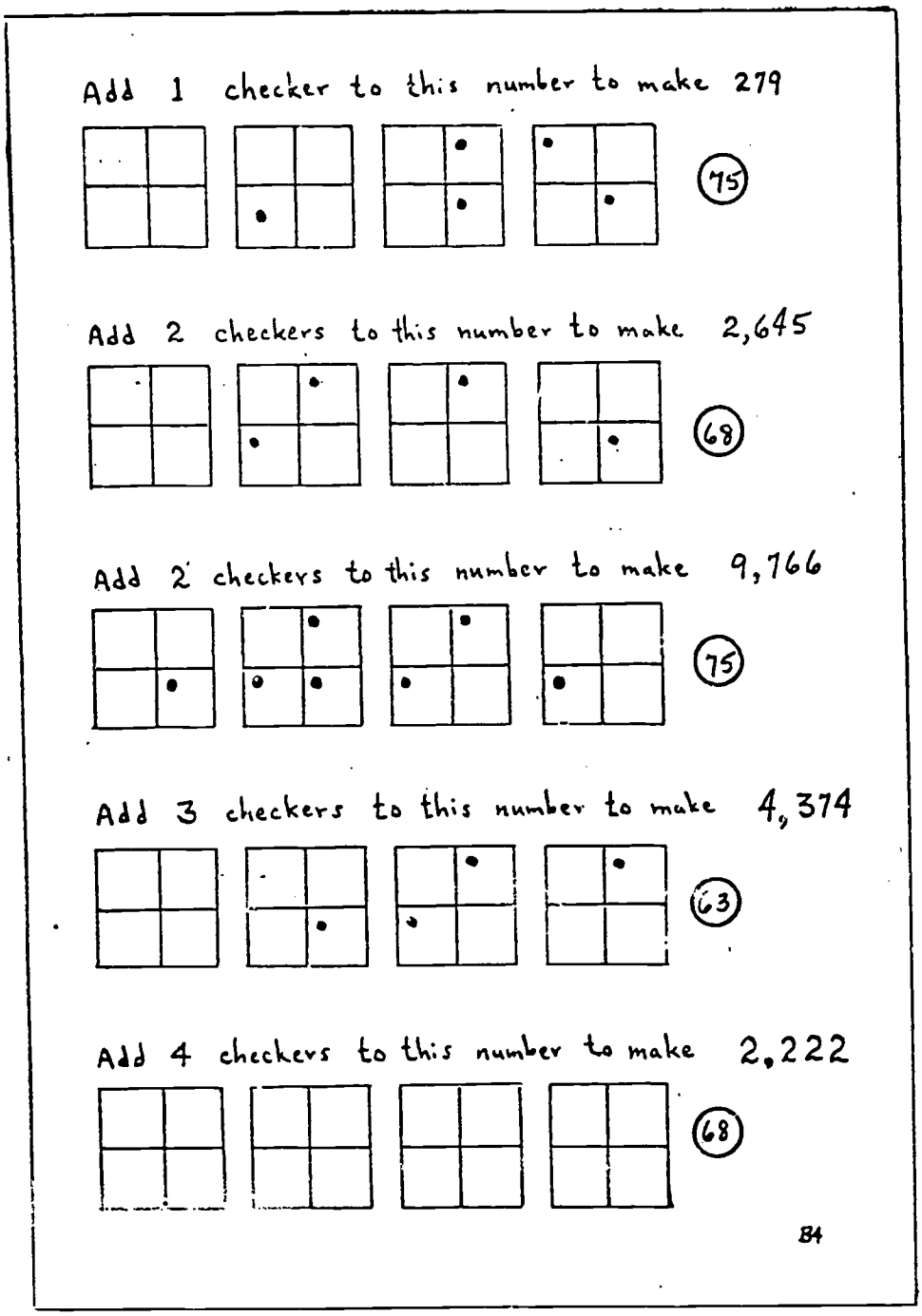
Third Grade, CSMP Test B (cont'd)



The only relatively common incorrect response was "158" (instead of "258") for the third item.

The correlation between this set of 5 items and the Kuhlmann Anderson test was .

Third Grade, CSMP Test B (cont'd)



There were no cases where students showed the required number on the Minicomputer by adding an incorrect number of checkers, but an average of 20% of the students were unable to show the required number though they did use the correct number of checkers.

The correlation between this set of items and the Kuhlmann-Anderson test was .49.

Third Grade, CSMP Test B (cont'd)

Write $>$, $=$ or $<$.

5	<	7	
$\hat{12}$		0	(58)
6		6	(92)
$\hat{6}$		$\hat{8}$	(41)
13		15	(93)
$\hat{12}+8$		$\hat{14}+8$	(34)
$\hat{2}$		1	(71)

Complete

$8+\hat{6} =$ (69)

$4+\hat{7} =$ (41)

$5+\hat{4} =$ (68)

$\hat{14}+3 =$ (37)

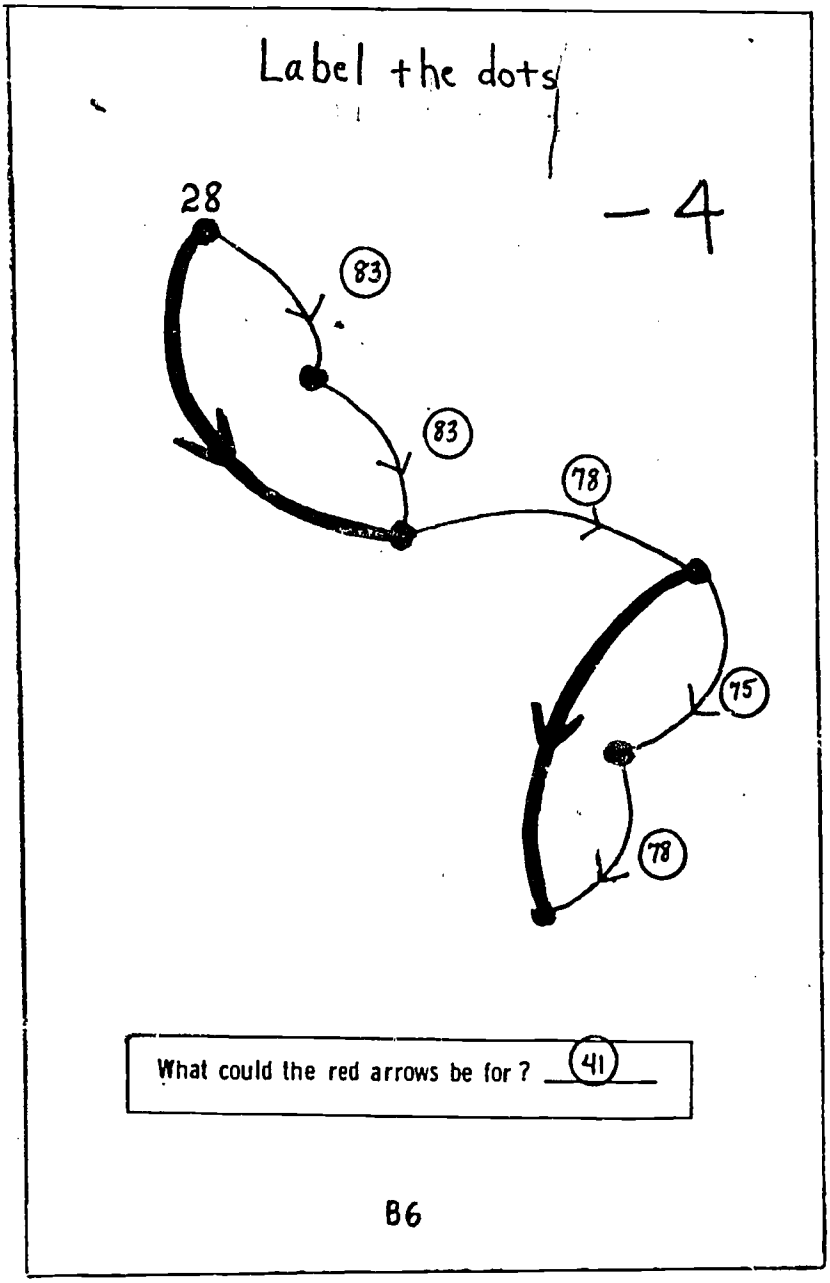
B5

The average percent correct for the 10 items was 60. The same set of items was given to second grade students (CSMP Test A) and the average percent correct was 50.

In the upper set of items, 15 percent of the students systematically reversed the inequality sign on all four items in which a negative number was present and another 15 percent reversed it for only those two items on which there was a negative number on both sides.

In the lower set of items, 15 percent of the students answered all 4 questions as if there were no "hats" present (giving a response of "14" to the calculation $8+\hat{6}$, for example). Twenty-five percent of the students gave a response of "3" instead of "$\hat{3}$" to "$4+\hat{7}$", and a response of "11" instead of "$\hat{11}$" to "$\hat{14}+3$".

Third Grade, CSMP Test B (cont'd)

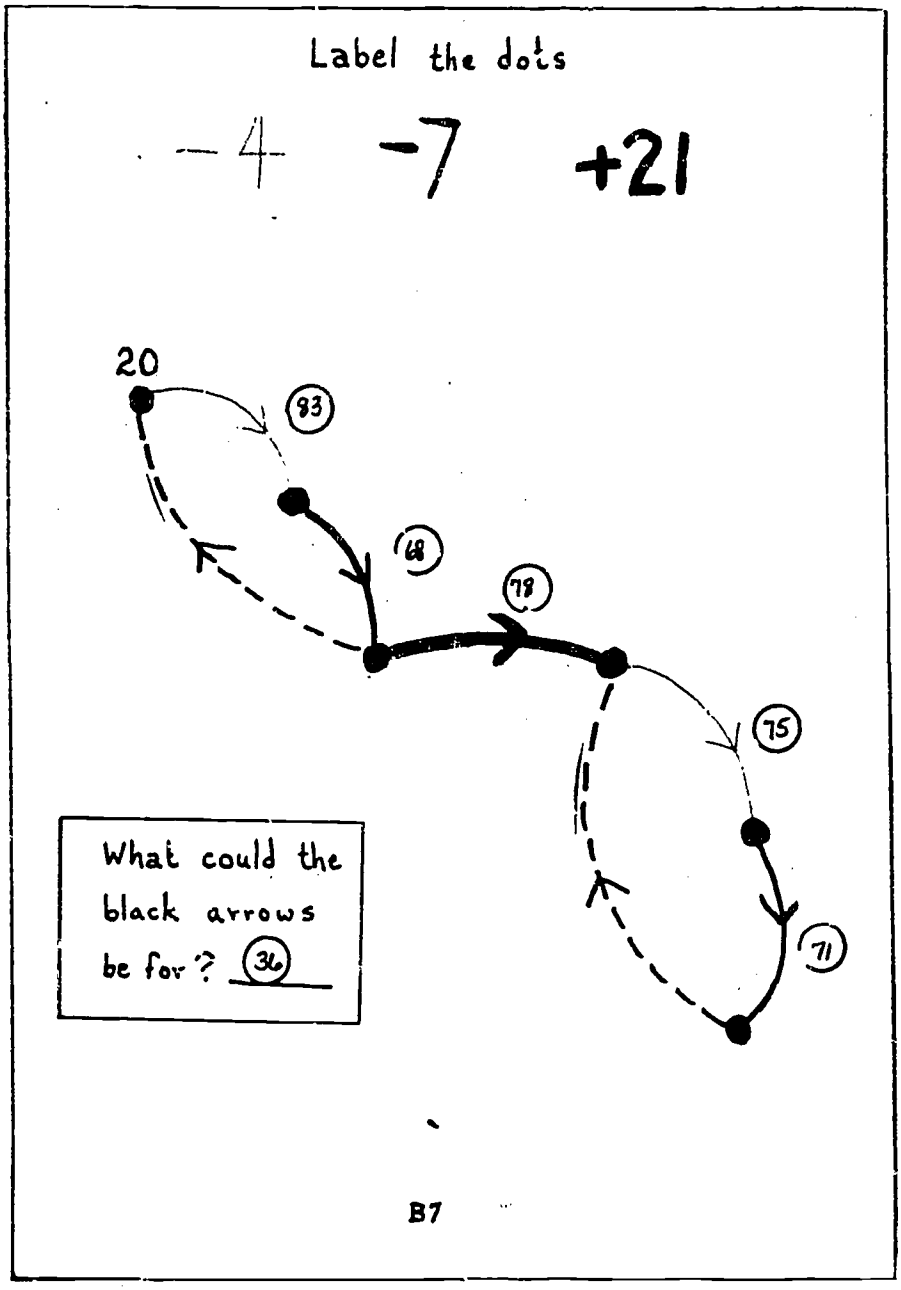


The mean percent correct for these 6 items was 73. The same page was given to second grade students (CSMP Test B) and the mean percent correct was 69.

Common incorrect responses to the last question, regarding the red arrows, were "+8", "+4", "5" and "+", though none were given by more than 5 percent of the students. (The heavy arrow was red on the original.)

The correlation of this set of 6 items with Kuhlmann Anderson test was .49.

Third Grade, CSMP Test B (cont'd)



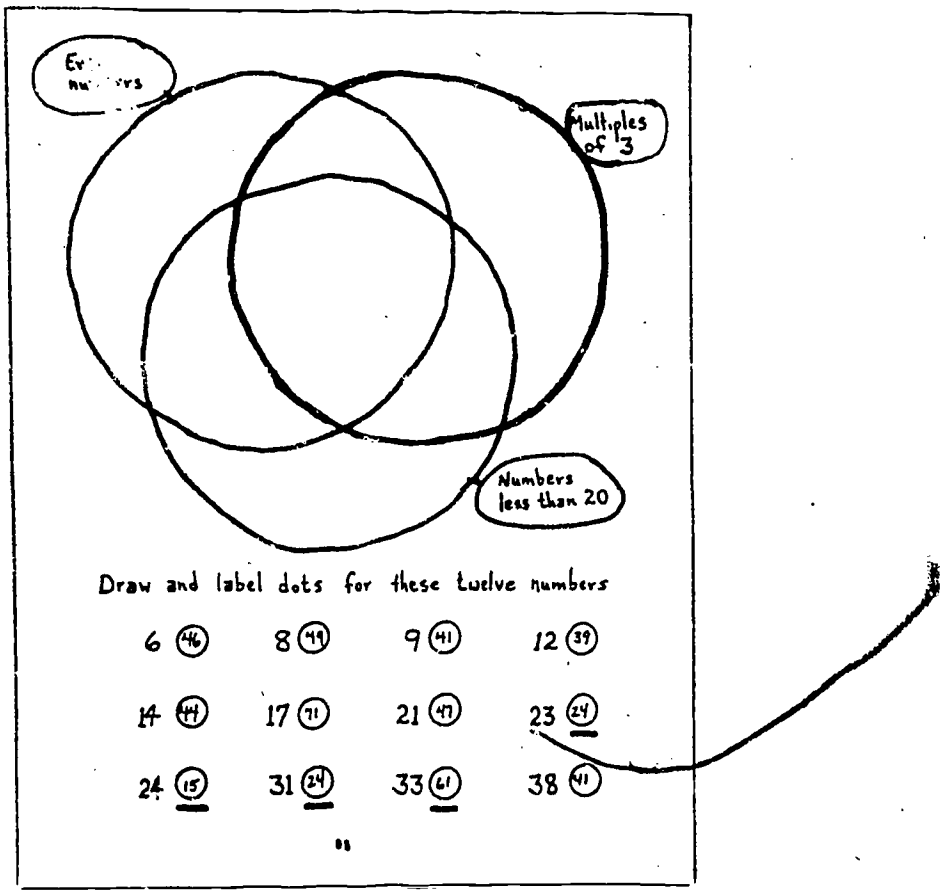
About 7 percent of the students evidently used the relations -4 or +21 when working with the arrow. These were the only instances of using the wrong one of the three given arrows.

Other than "+11", the most common answers to the last question, regarding the black arrows, were "11" (without the plus sign), "8" and "$\frac{1}{2}x$". (The dotted arrows were originally in black.)

The correlation of this set of items and the Kuhlmann-Anderson test was .65.

Third Grade, CSMP Test B (cont'd)

"The blue string is for even numbers. The yellow string is for numbers which are multiples of 3. And the red string is for numbers less than 20. Below the picture it says, 'Draw and label dots for these twelve numbers.' You can see the numbers at the bottom of the page. For instance you have to find where the dot for 6 goes and put the dot where it belongs in the string picture and label it. Then do the same for the other numbers." (The strings were appropriately in color on the original.)



For each number, the percent of students who placed a dot for that number in the correct region is circled beside the number. That circled percent is underlined when more than 10 percent of the students did not respond to the corresponding number. (The mean percent omits, for all items, was 10.)

The mean percent correct across items was 42. The hardest numbers to place correctly were "23" and "31", which were outside all the given strings and, surprisingly, "24", the only number inside the blue and yellow strings and outside the red string. The easiest number to place was "17", which was, in a sense, functionally opposite to "24" (i.e., outside the blue and yellow strings and inside the red string).

The correlation between this set of 12 items and the Kuhlmann Anderson test was .52.

Appendix G

MANS Tests A and B

Each of the MANS tests was composed of five subtests. Separate analyses were performed for each subtest, and the data from these analyses are presented in the following order:

MANS - Test A: Subtest 1, Sequences
Subtest 2, Equation Fluency
Subtest 3, Functions
Subtest 4, Number Line Estimation
Subtest 5, Computation I

MANS - Test B: Subtest 1, Labelling Number Lines
Subtest 2, Sentences About 8
Subtest 3, Word Problems
Subtest 4, Number Sentence Pictures
Subtest 5, Computation II

For each subtest, three kinds of information are presented: a) item analysis data, b) summary subtest data and c) class mean data.

a) Item Analysis Data
For each actual test page, a page appears in the Appendix with the following information:
i) At the top of the page, a shortened version of the actual directions is given. Sample problems were discussed with the class and ample demonstrations were given. The tester circulated around the room to answer questions and to check on the students' understanding of the directions. Although there was a time limit for each subtest, only one subtest was speeded (Number Line Estimation in which 2½ minutes was allowed for the 13 items). The total working time for MANS A and MANS B was 22 minutes. Time limits for each subtest were determined following pre-testing research and allowed almost all students to complete the subtests.
ii) In the middle of the page, a reduced version of the actual test page is given. Next to each test item, the mean percent correct for all second grade CSMP students who took the test (n=337 for Test A, 352 for Test B) is given by a circled entry (Ⓧⓧ). Following this entry, is the mean percent correct for all second grade Non-CSMP students who took the test (n=324 for Test A and 348 for Test B). This entry is boxed (🄾🄾).

b) Summary Subtest Data
At the bottom of the page, the mean scores on the subtest are given for all CSMP students and for all Non-CSMP students. In addition, for each of these two groups of students, mean scores by ability group are given. For example, a mean score is given for all CSMP students in quintile 1, i.e. for students with a mean score on the covariate test (reading or general ability) of less than 20. Similarly mean scores are given for CSMP Q2, Q3, Q4 and Q5 and for the various Non-CSMP quintiles. The distribution of students across quintiles is shown below. It can be seen that there are somewhat

higher percentages of CSMP students in the upper quintiles; within a given quintile, however, the mean covariate scores were nearly equal.

	MANS A		MANS B	
	CSMP	Non-CSMP	CSMP	Non-CSMP
Quintile 1	13%	15%	10%	15%
Quintile 2	16%	16%	12%	16%
Quintile 3	25%	27%	26%	29%
Quintile 4	26%	24%	29%	24%
Quintile 5	20%	18%	23%	16%
All Students Combined	100%	100%	100%	100%

Also given in the summary subtest data (bottom of each item analysis page) is the reliability (KR20) for the subtest and correlations with various other tests (CSMP A, CSMP B, Kuhlmann-Anderson, and standardized math tests).

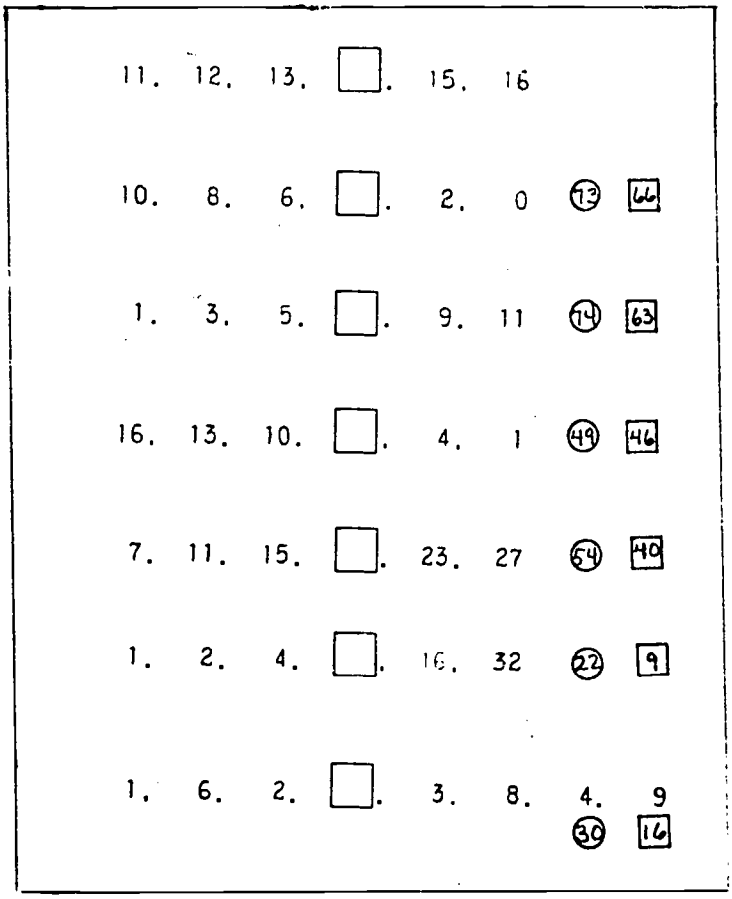
c) Class Mean Data
The mean score on the subtest, for each class which took the subtest, is plotted on a graph against the mean score on the ability or reading score for that class. Through the resulting set of points, one for each class, the regression line of test score on covariate has been drawn. These graphs provide a visual comparison between CSMP and Non-CSMP classes and the reader can also see the degree to which class scores are predicted by the covariate. In the corner of each graph, summary statistical data from the Analysis of Covariance with blocking on sites* is given, including adjusted means and p-values.

*Blocking on sites removes any systematic variation from site-to-site from the unexplained error variance. This model assumes there are four main sources contributing to the variance between scores:
a) due to entering ability as measured by the covariate (1 degree of freedom)
b) due to systematic differences between the four sites (3 degrees of freedom)
c) due to the differences in curricula (1 degree of freedom)
d) due to random unexplained error (22 (MANS A) and 24 (MANS B) degrees of freedom)

The F-test compares variances c) and d); if c) is proportionally much larger than d), it is unlikely that differences in curricula are zero and produce only random unexplained error. Then the resulting p-value is low and when below .05 one says the differences are "significant" - unlikely to have been produced by chance alone.

MANS A, Subtest 1: Sequences

"Look at the first row of numbers on this page. The numbers in the first row are 11, 12, 13, then a box, then 15, and then 16. The numbers are in a certain order and we have to figure out what the pattern is so that we can put the right number in the box." (The students were given time to work on this problem and the answer was explained.) "Now when I tell you to start you do the rest of the questions on this page but be careful because each pattern is different."



Means

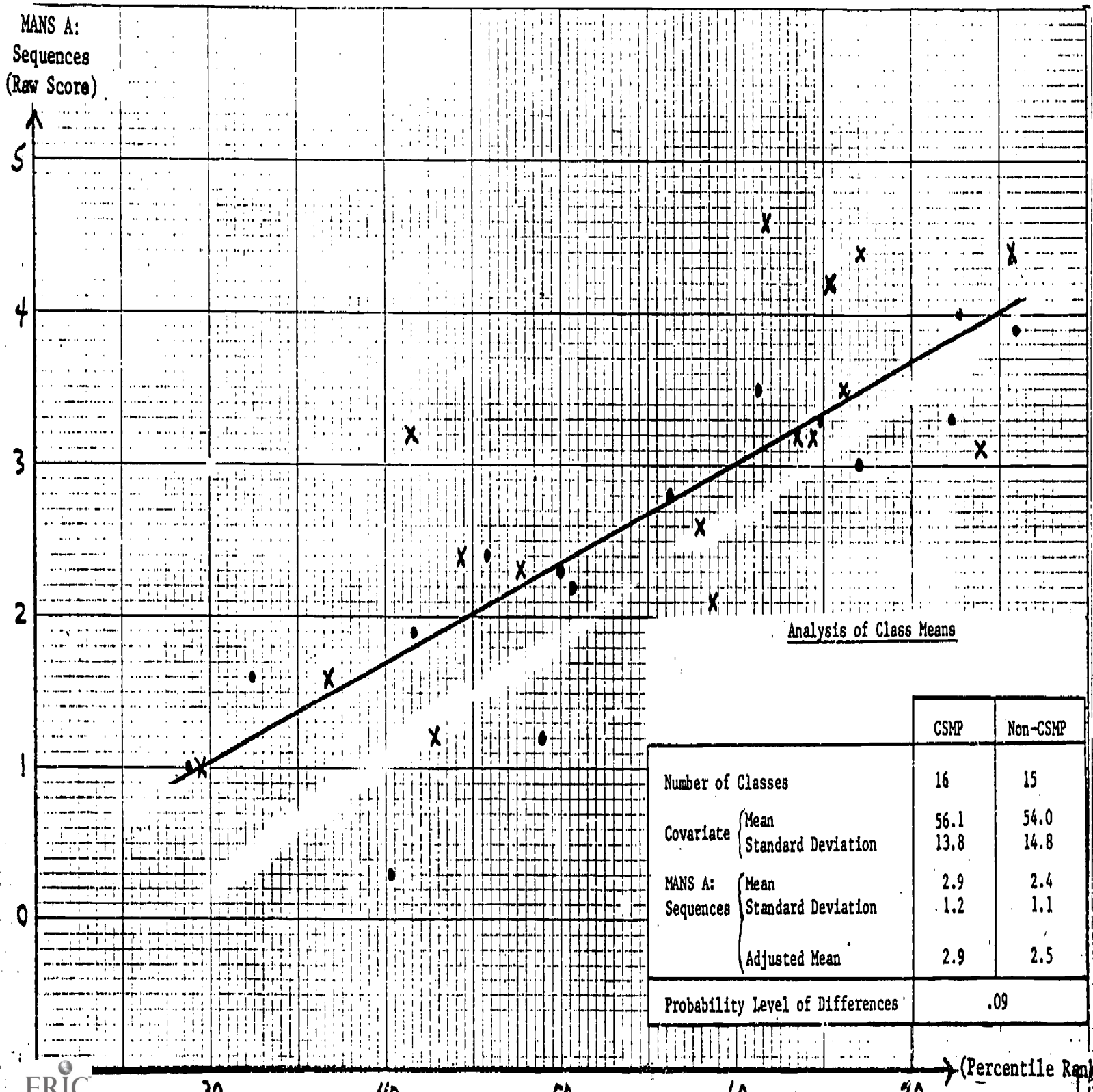
	CSMP	Non-CSMP
All Students Combined	2.9	2.4
Quintile 1	1.9	1.5
Quintile 2	1.9	1.1
Quintile 3	2.1	2.0
Quintile 4	3.7	3.0
Quintile 5	4.2	3.8

Correlations (Subtest 1 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.80	.77
CSMP-Specific Test A	.61	
CSMP-Specific Test B	.53	
Kuhlmann-Anderson	.63	.63
Reading Tests[1]	.43	.42
Math Tests[1]	.61	.61

[1]Mean correlation from three different standardized reading (math) tests administe at various sites.

MANS Test A: Sequences versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

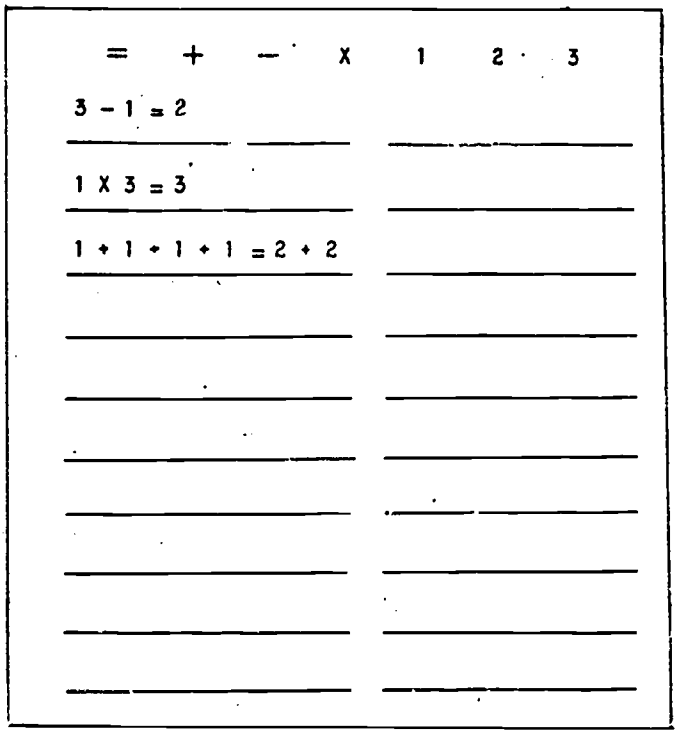
		CSMP	Non-CSMP
Number of Classes		16	15
Covariate	Mean	56.1	54.0
	Standard Deviation	13.8	14.8
MANS A: Sequences	Mean	2.9	2.4
	Standard Deviation	1.2	1.1
	Adjusted Mean	2.9	2.5
Probability Level of Differences		.09	

MANS A, Subtest 2: Equation Fluency

"Look at the numbers and things at the top of the page" (the signs and numbers were described to the students). We want to write as many true number sentences as we can using only these signs and numbers. You can use each sign or number as many times as you want in the same sentence. And you don't have to use them all in the same sentence."

(The three sentences given as examples were explained.)

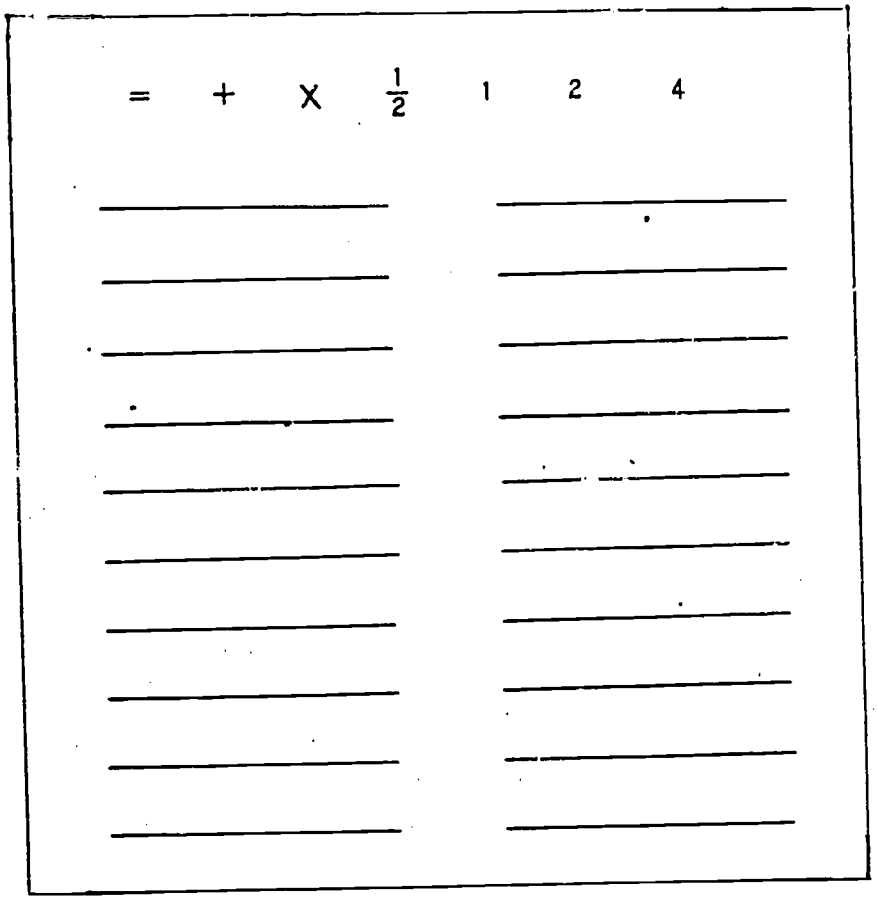
"Now when I tell you to start, I want you to write as many other number sentences as you can. Work as fast as you can until I tell you to stop."



	CSMP			Non-CSMP		
	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences
All Students Combined	3.7	0.6	3.1	3.6	1.0	2.6
Quintile 1	2.2	1.1	1.1	2.4	1.3	1.1
Quintile 2	2.9	0.6	2.3	2.2	1.2	1.0
Quintile 3	3.2	0.7	2.5	3.3	1.3	2.0
Quintile 4	4.6	0.5	4.1	4.2	0.7	3.5
Quintile 5	4.8	0.4	4.4	5.7	0.4	5.3

MANS A, Subtest 2: Equation Fluency (cont.)

"This problem is just like the last page. I want you to write as many true number sentences as you can. Work as fast as you can until I tell you to stop."



	CSMP			Non-CSMP		
	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences
All Students Combined	4.3	0.8	3.5	3.4	1.2	2.2
Quintile 1	2.6	1.7	0.9	2.1	1.6	0.5
Quintile 2	3.1	0.8	2.3	2.2	1.1	1.1
Quintile 3	3.4	1.0	2.4	3.2	1.6	1.6
Quintile 4	5.3	0.5	4.8	4.3	0.8	3.5
Quintile 5	5.9	0.4	5.5	5.0	0.7	4.3

MANS A, Subtest 2: Equation Fluency (cont.)

Total Score (Both Pages)

	CSMP			Non-CSMP		
	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences
All Students Combined	8.0	1.4	6.6	7.0	2.2	4.8
Quintile 1	4.8	2.8	2.0	4.5	2.9	1.6
Quintile 2	6.0	1.4	4.6	4.4	2.3	2.1
Quintile 3	6.6	1.7	4.9	6.5	2.9	3.6
Quintile 4	9.9	1.0	8.9	8.5	1.5	7.0
Quintile 5	10.7	0.8	9.9	10.7	1.1	9.6

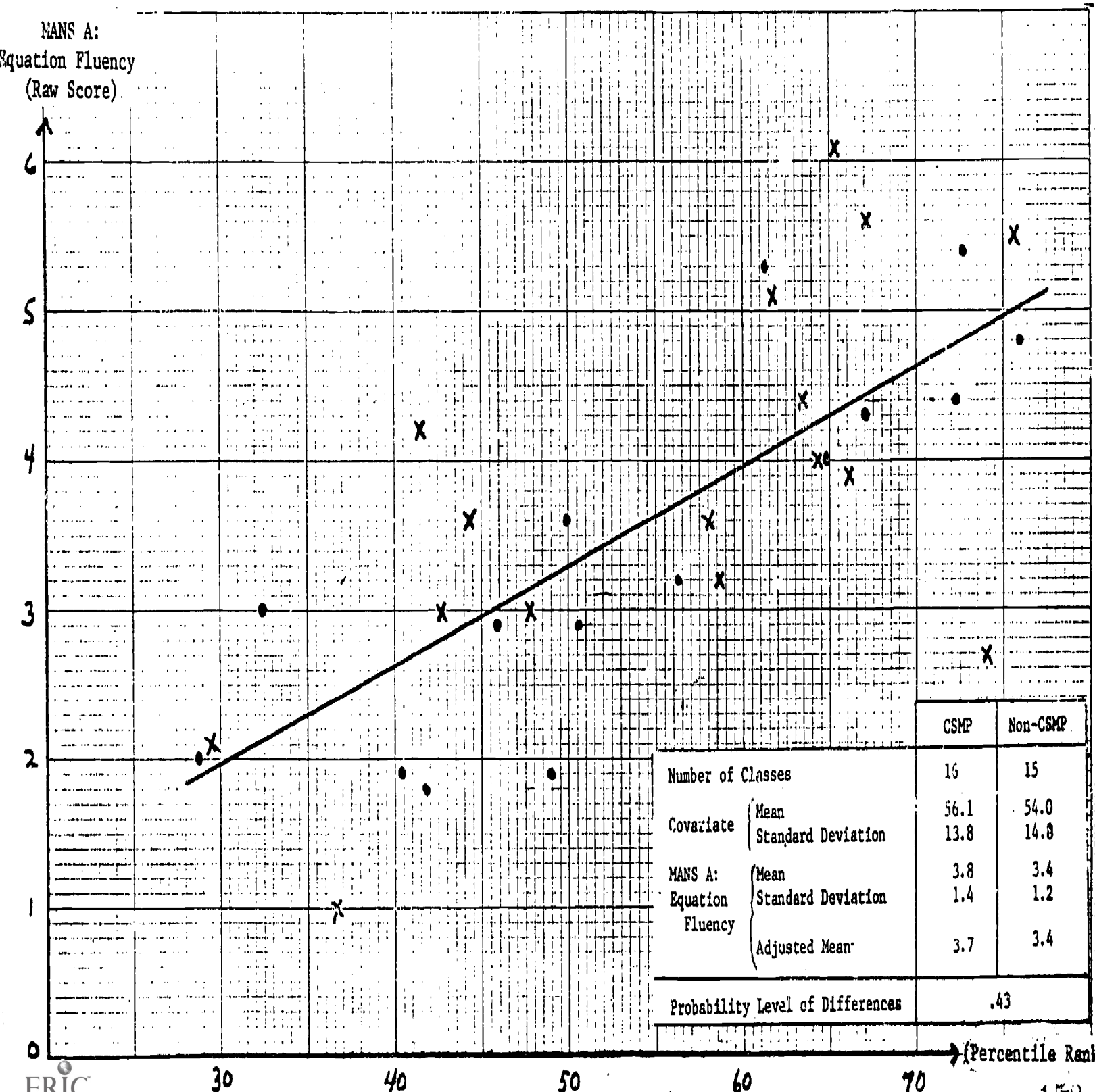
*For purposes of class mean computation, the total number of true sentences was multiplied by one-half and tests of significance were based on these numbers. Hindsight indicates this was not the best number to use. A more informative score (and one which would have been more likely to disclose significant differences) would have been "total number of true sentences minus total number of false sentences."

Correlations (Subtest 2 v Other Tests)

	CSMP	Non-CSMP
CSMP-Specific Test A	.60	
CSMP-Specific Test B	.52	
Kuhlmann-Anderson	.67	.67
Reading Tests[1]	.36	.43
Math Tests[1]	.53	.56

[1]Mean correlation from three different standardized reading (math) tests administered at various sites.

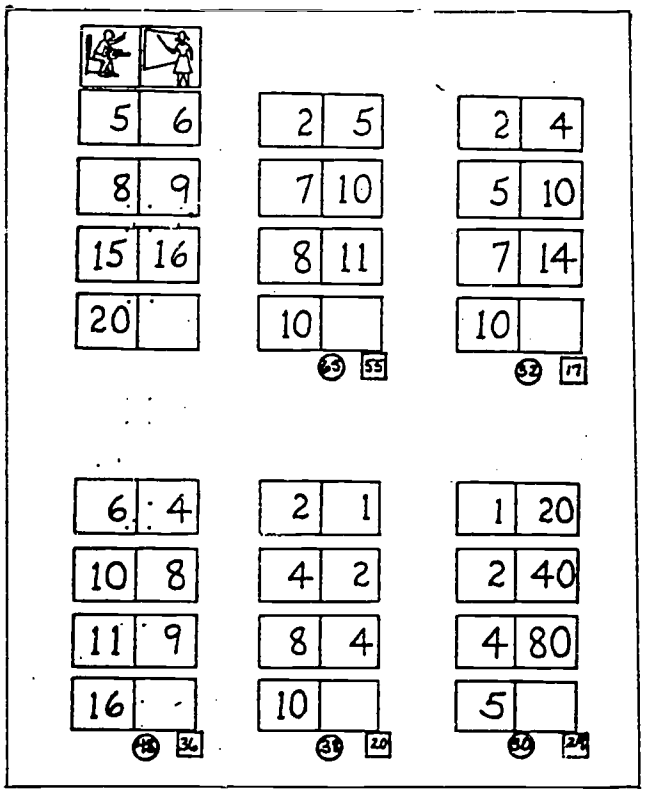
MANS Test A: Equation Fluency versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



		CSMP	Non-CSMP
Number of Classes		16	15
Covariate	Mean	56.1	54.0
	Standard Deviation	13.8	14.8
MANS A: Equation Fluency	Mean	3.8	3.4
	Standard Deviation	1.4	1.2
	Adjusted Mean	3.7	3.4
Probability Level of Differences		.43	

MANS A, Subtest 3: Functions

"Look at the pictures at the top of the page. There is a picture of a student at his desk. Then there is a picture of a teacher at the blackboard. Look at the numbers in the boxes below these pictures. There's a 5 and beside it there's a 6. That means a student gave her a 5 and she did something to it and got 6. The next two numbers are 8 and 9. A student gave her 8, she did the same thing to it and got 9. Then the next two numbers are 15 and 16. Then there is a 20 and an empty box. Can you figure out what could go in the empty box?" (The students were given time to work on this problem and the answer was explained.) "The teacher is adding one <u>each</u> time; 5+1=6, 8+1=9; 15+1=16, 20+1=21. Now, there are 5 other problems on the page like this one. For each one, you have to figure out what it was that the teacher was doing to the numbers, and then put the right number in the empty box. Be careful because she's doing a <u>different</u> thing for each of the five problems. She won't be adding one to any of the others."



<u>Means</u>

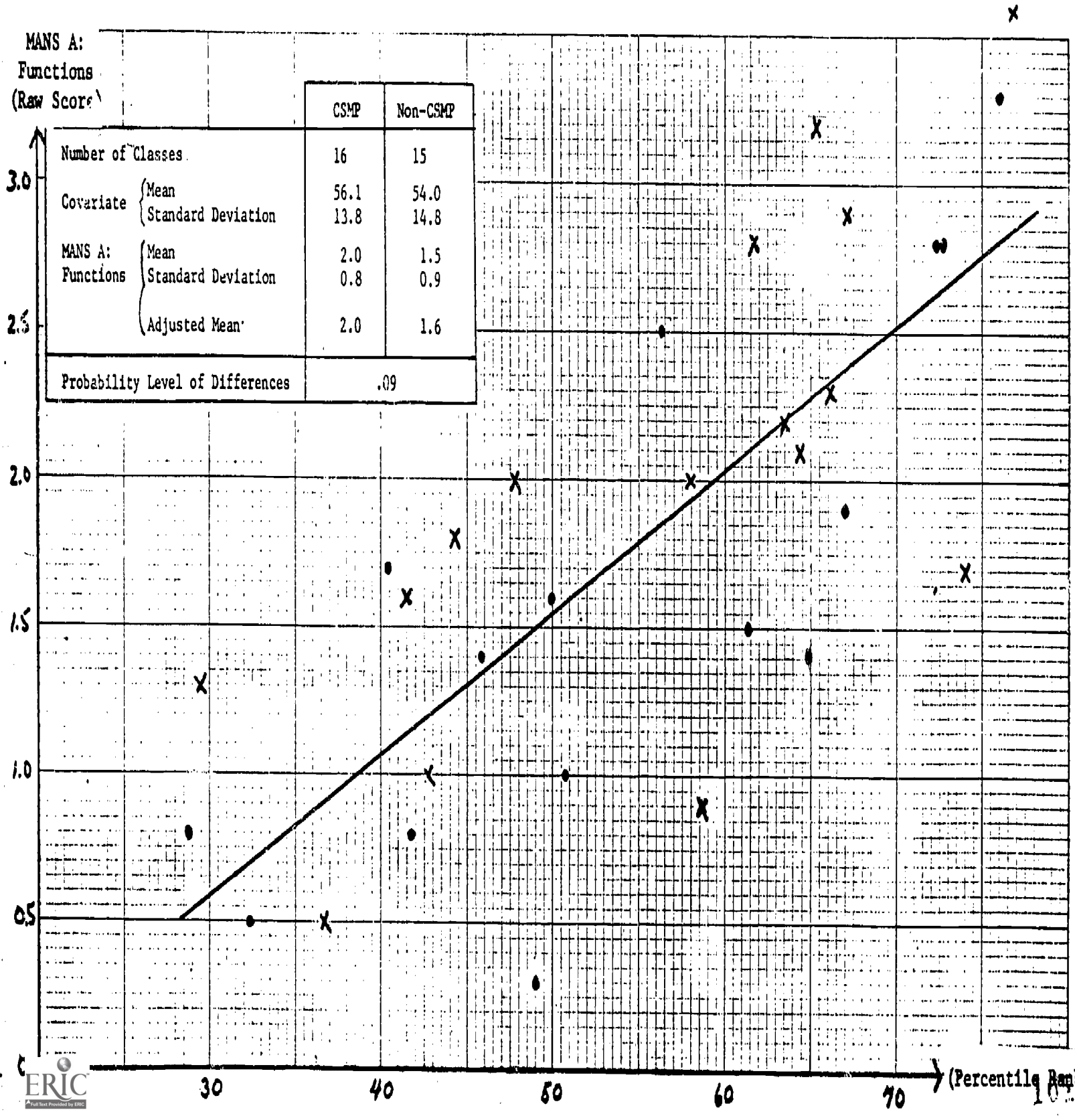
	CSMP	Non-CSMP
All Students Combined	2.0	1.5
Quintile 1	0.7	0.7
Quintile 2	1.4	0.9
Quintile 3	1.4	1.1
Quintile 4	2.6	1.9
Quintile 5	3.0	2.7

<u>Correlations</u> (Subtest 3 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.73	.66
CSMP-Specific Test A	.54	
CSMP-Specific Test B	.57	
Kuhlmann-Anderson	.66	.62
Reading Tests[1]	.41	.39
Math Tests[1]	.52	.53

[1]Mean correlation from three different standardized reading (math) tests administered at various sites.

MANS Test A: Functions versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



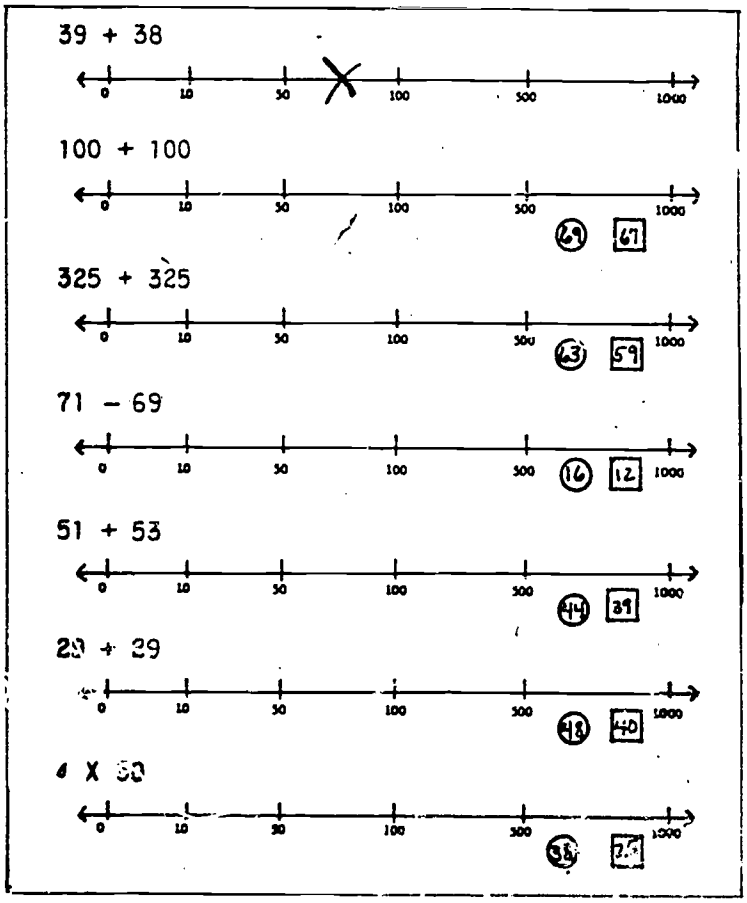
		CSMP	Non-CSMP
Number of Classes		16	15
Covariate	Mean	56.1	54.0
	Standard Deviation	13.8	14.8
MANS A: Functions	Mean	2.0	1.5
	Standard Deviation	0.8	0.9
	Adjusted Mean	2.0	1.6
Probability Level of Differences		.09	

MANS A, Subtest 4: Number Line Estimation

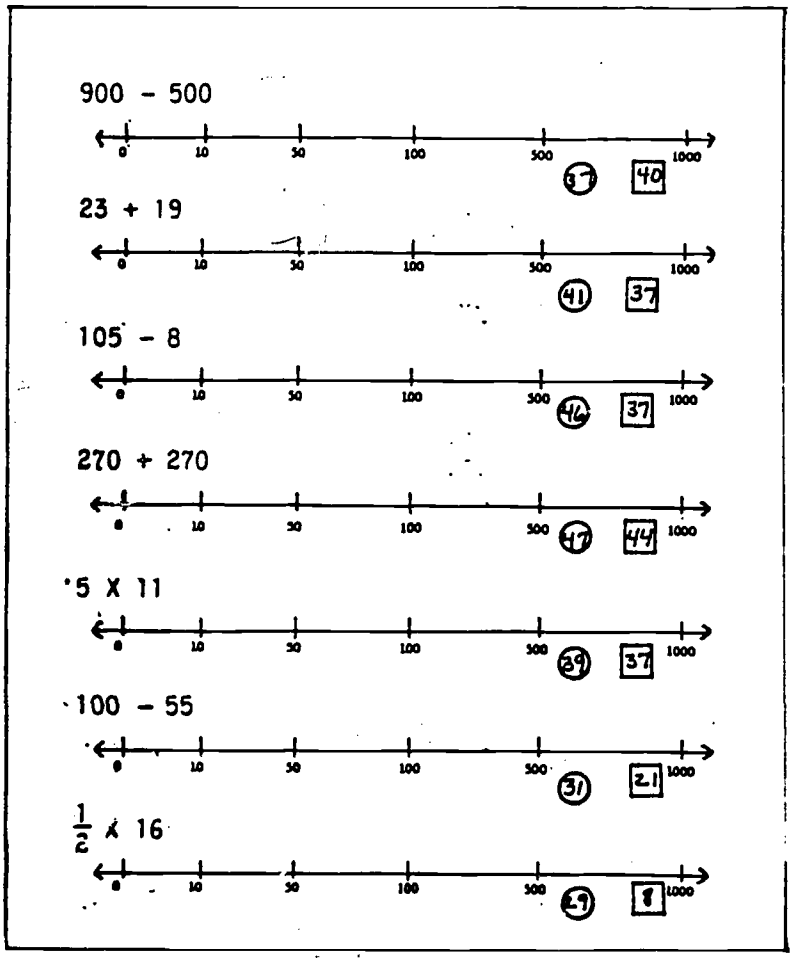
"Look at this funny looking number line. On the next two pages there will be a lot of number problems. For each one you will have to figure out where the answer would go on the number line. It could go here between 0 and 10, here between 10 and 50," (etc. for successive intervals). "You just put a big x on the number line to show which two numbers the answer goes between. Now there are a whole lot of problems and you won't have much time to do them. You should not try to calculate the exact answer; just decide quickly where the answer would probably go on the number line and put a big x on the right place on the number line."

(The students were given time to do the sample problem (39+38) on a separate sheet of paper, and the answer was explained. The fact that they had only 10 seconds to do each problem was demonstrated.)

"Now open your books to this page. On the next two pages you will have a whole lot to do the same way. I'll give you two and a half minutes to do them so you'll have to work fast. I'll let you know when half the time is up so you can see how you're doing."



MANS A, Subtest 4: Number Line Estimation (cont.)



Means

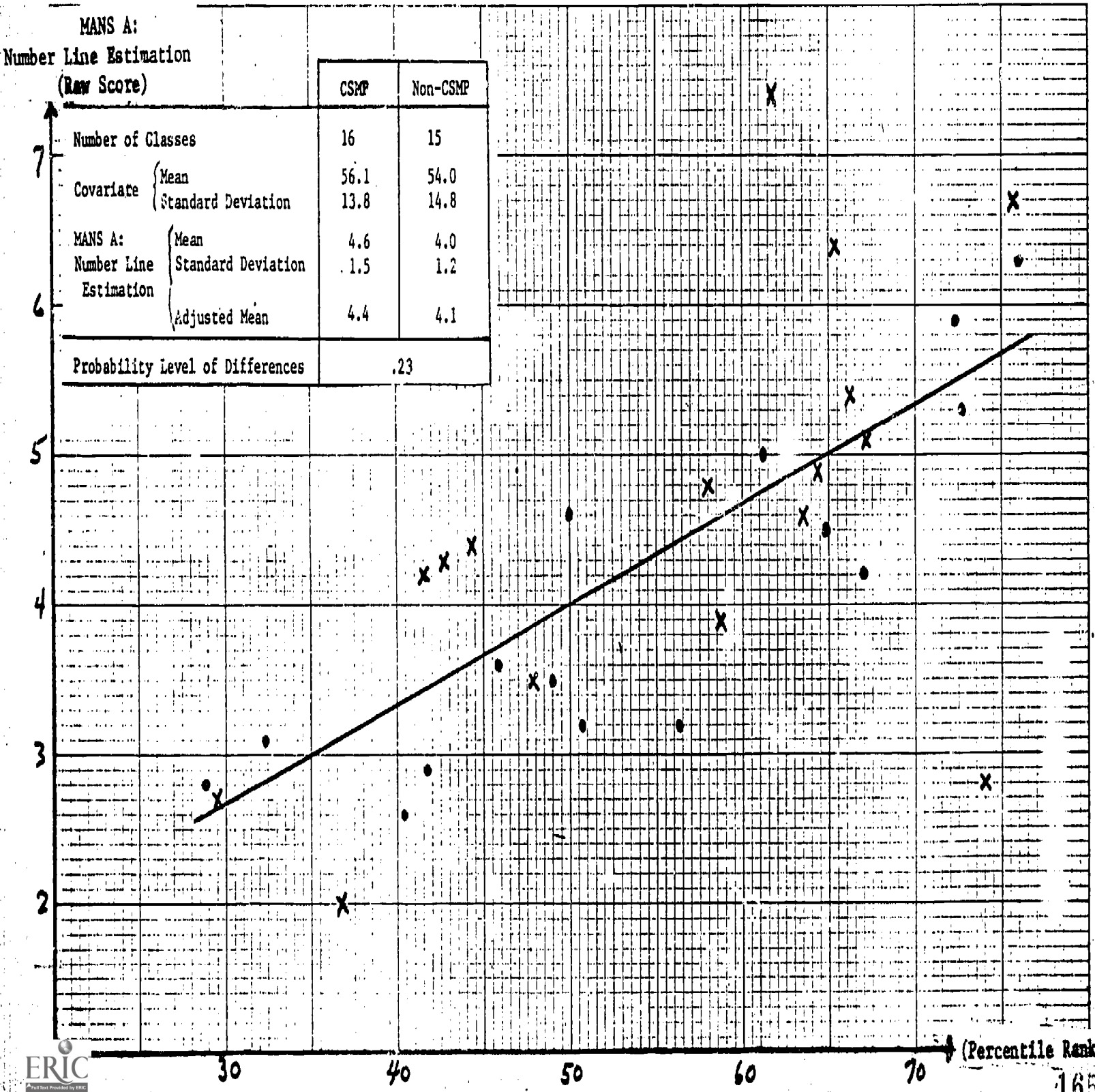
	CSMP	Non-CSMP
All Students Combined	4.5	4.0
Quintile 1	3.1	2.8
Quintile 2	3.2	3.1
Quintile 3	3.6	3.5
Quintile 4	5.6	4.2
Quintile 5	6.1	6.0

Correlations (Subtest 4 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.75	.60
CSMP-Specific Test A	.54	
CSMP-Specitif Test B	.50	
Kuhlmann-Anderson	.65	.60
Reading Tests[1]	.28	.34
Math Tests[1]	.61	.61

[1]Mean correlations from three different standardized reading (math) tests administered at various sites.

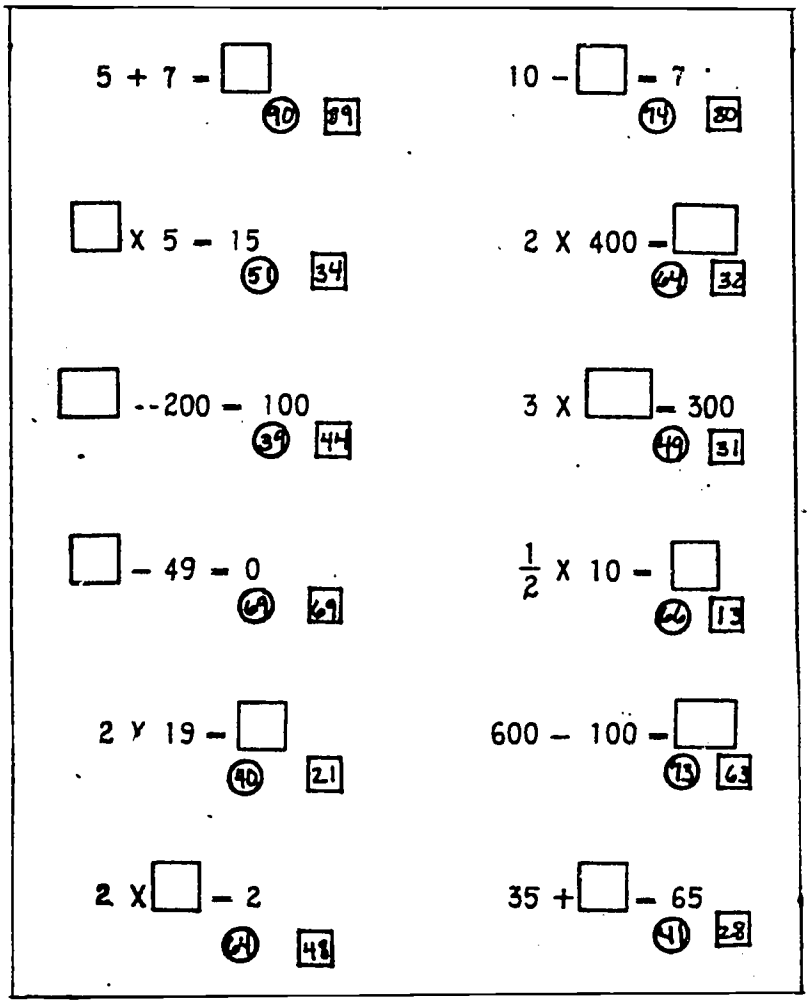
MANS Test A: Number Line Estimation versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



		CSMP	Non-CSMP
Number of Classes		16	15
Covariate	Mean	56.1	54.0
	Standard Deviation	13.8	14.8
MANS A: Number Line Estimation	Mean	4.6	4.0
	Standard Deviation	1.5	1.2
	Adjusted Mean	4.4	4.1
Probability Level of Differences		.23	

MANS A, Subtest 5: Computation I

"You have to figure out what goes in the boxes to make the number sentences true."



Means

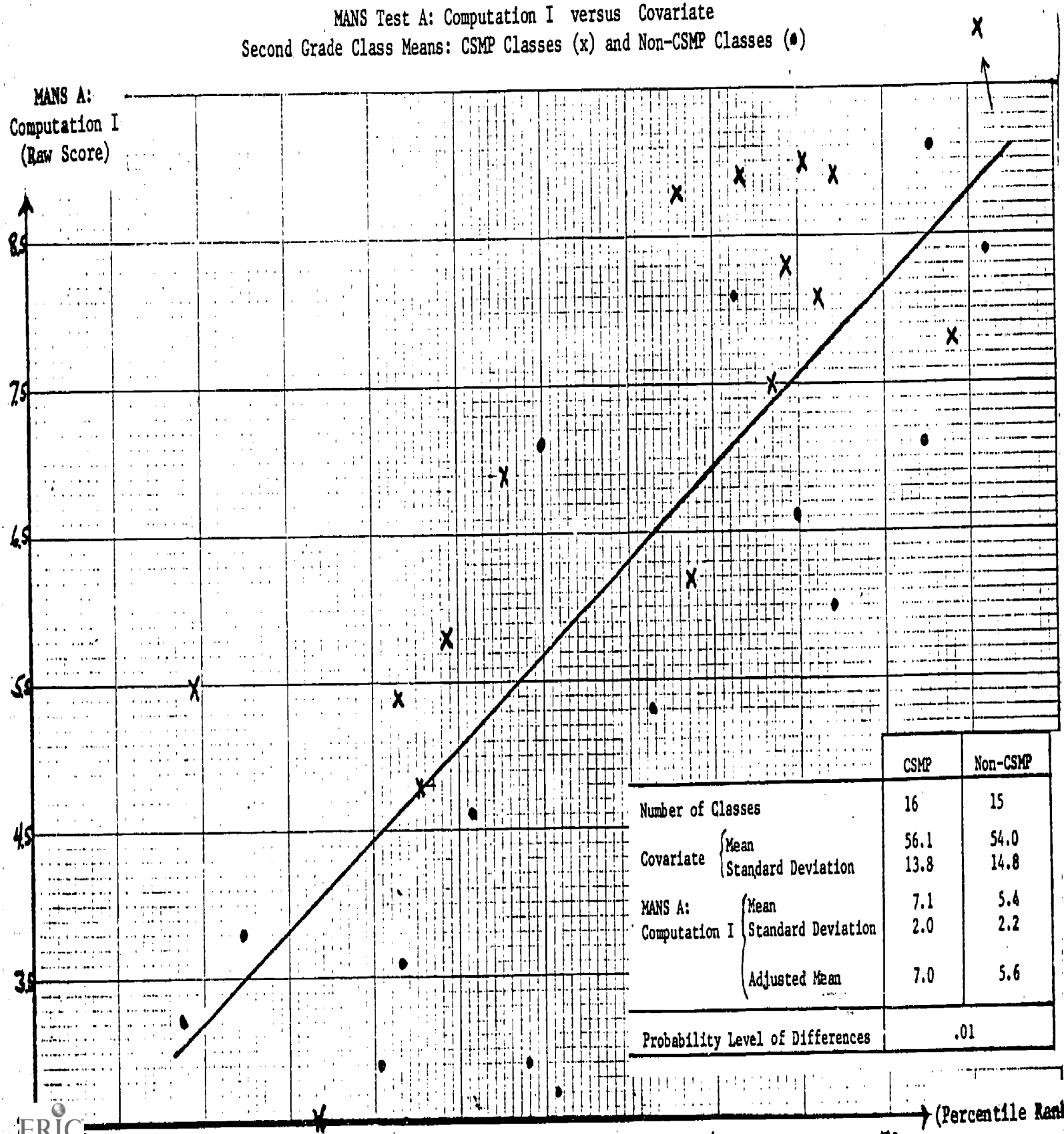
	CSMP	Non-CSMP
All Students Combined	7.0	5.3
Quintile 1	3.3	3.5
Quintile 2	5.5	3.5
Quintile 3	5.9	4.4
Quintile 4	8.8	6.3
Quintile 5	9.7	8.5

Correlations (Subtest 5 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.86	.84
CSMP-Specific Test A	.73	
CSMP-Specific Test B	.71	
Kuhlmann-Anderson	.68	.71
Reading Tests[1]	.59	.47
Mean Tests[1]	.73	.69

[1]Mean correlation from three different standardized reading (math) tests administered at various sites.

MANS Test A: Computation I versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



		CSMP	Non-CSMP
Number of Classes		16	15
Covariate	Mean	56.1	54.0
	Standard Deviation	13.8	14.8
MANS A: Computation I	Mean	7.1	5.4
	Standard Deviation	2.0	2.2
	Adjusted Mean	7.0	5.6
Probability Level of Differences		.01	

MANS Test A, Intercorrelations
(First Entry: CSMP, Second Entry: Non-CSMP)[1]

	Subtest 1		Subtest 2		Subtest 3		Subtest 4	
Subtest 1: Sequences								
Subtest 2: Equation Fluency	.54	.53						
Subtest 3: Functions	.56	.54	.52	.49				
Subtest 4: Number Line Estimation	.48	.44	.53	.45	.44	.40		
Subtest 5: Computation I	.62	.64	.64	.65	.63	.56	.56	.57

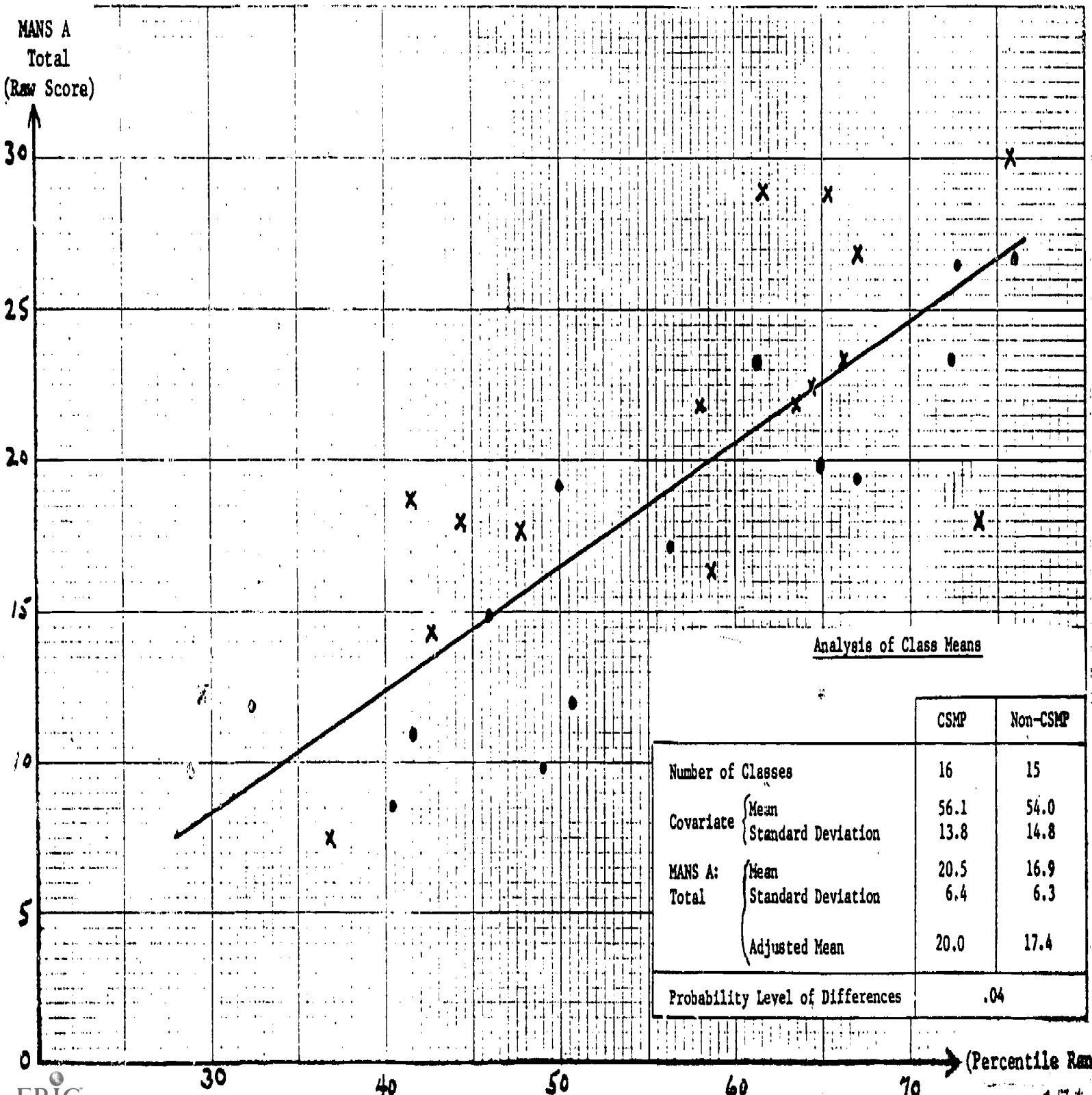
[1]Based on 426 CSMP students and 390 Non-CSMP students.

Correlations Between Total Score, MANS A and Other Tests

	CSMP Students[1]		Non-CSMP Students[1]	
Correlation with: CSMP-Specific Test A	.77	(294)		
CSMP-Specific Test B	.73	(143)		
Kuhlmann-Anderson	.78	(129)	.79	(114)
Gates-MacGinitie Reading	.62	(77)	.52	(69)
SAT Reading	.60	(83)	.63	(90)
CPT Reading	.47	(49)	.53	(50)
CTBS Math	.84	(235)	.84	(211)
SAT Math	.67	(83)	.75	(90)
CPT Math	.73	(49)	.68	(50)

[1]Number in parenthesis is number of students on whom correlation is based.

MANS Test A: Total Score versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)

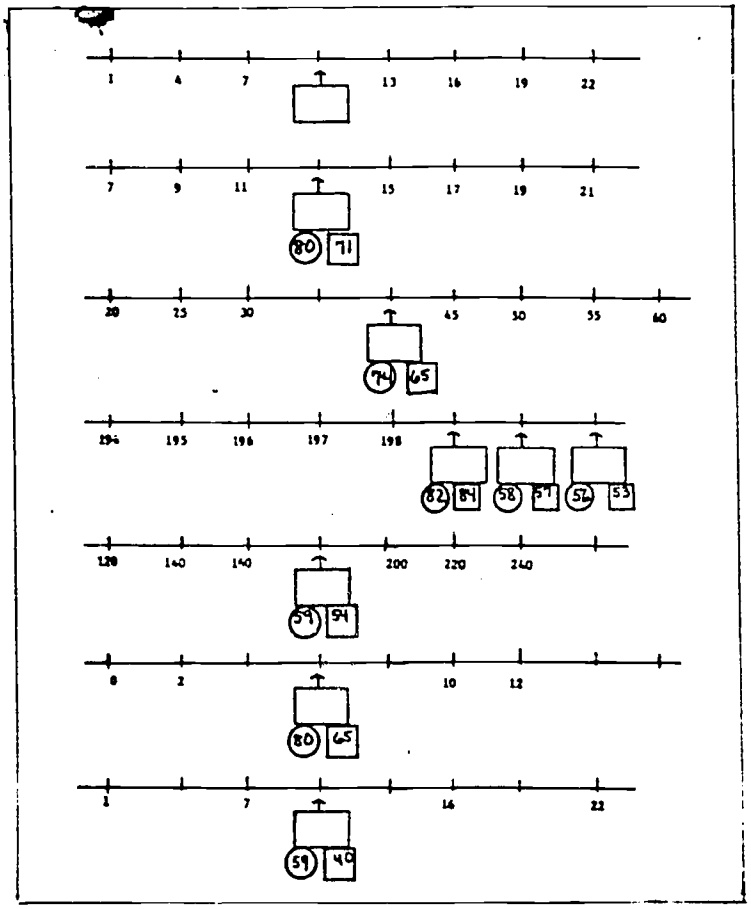


Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		16	15
Covariate	Mean	56.1	54.0
	Standard Deviation	13.8	14.8
MANS A: Total	Mean	20.5	16.9
	Standard Deviation	6.4	6.3
	Adjusted Mean	20.0	17.4
Probability Level of Differences		.04	

MANS B, Subtest 1; Labelling Number Lines

"Look at the top row of numbers on this page. The numbers on that top number line are 1, 4, 7, then a box, then 13, 16, 19 and 22. You have to figure out what number would go in the box." (The students were given time to work on this problem and the answer was explained.) "'Ten' is the right answer because the numbers are going up by three each time; each mark is three more than the one before it. When I tell you to start you do the rest of the questions on this page."



Means

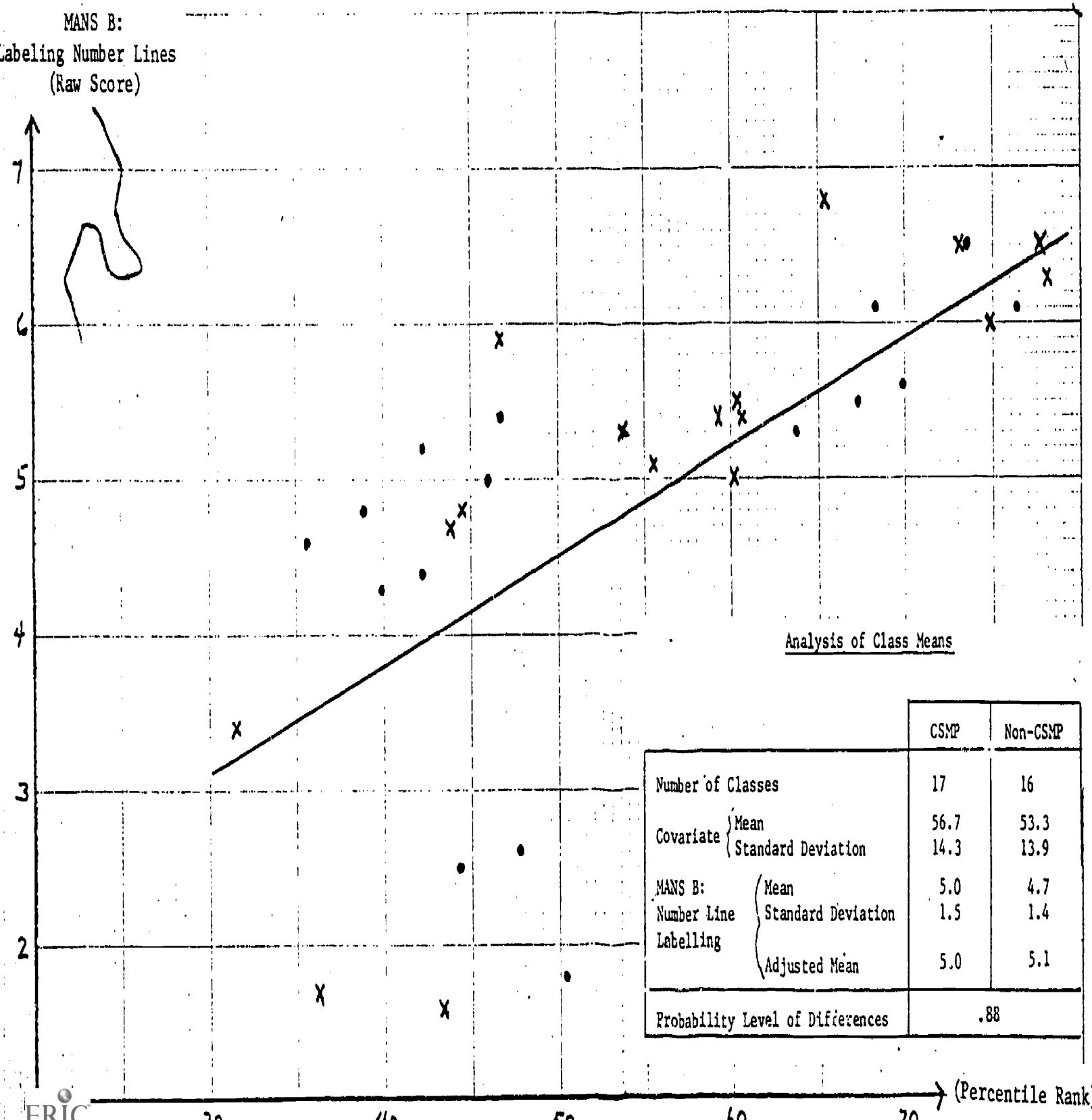
	CSMP	Non-CSMP
All Students Combined	5.1	4.7
Quintile 1	3.1	2.6
Quintile 2	3.3	3.9
Quintile 3	4.4	4.1
Quintile 4	5.9	5.5
Quintile 5	6.9	6.3

Correlations (Subtest 1 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.85	.79
CSMP-Specific Test A	.73	
CSMP-Specific Test B	.46	
Kuhlmann-Anderson	.63	.56
Reading Tests[1]	.48	.42
Math Tests[1]	.68	.63

[1]Mean correlation from three different standardized reading (math) tests administered at various sites.

MANS Test B: Labeling Number Lines versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)

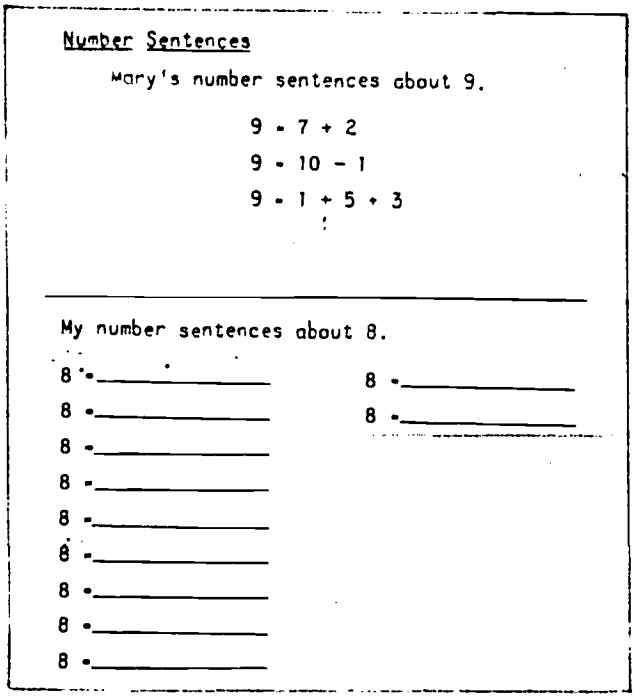


Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		17	16
Covariate	Mean	56.7	53.3
	Standard Deviation	14.3	13.9
MANS B: Number Line Labelling	Mean	5.0	4.7
	Standard Deviation	1.5	1.4
	Adjusted Mean	5.0	5.1
Probability Level of Differences		.88	

MANS B, Subtest 2: Sentences About 8

"Look at the top of this page. It says: 'Number sentences. Mary's number sentences about 9. 9=7+2, 9=10-1, 9=1+5+3.' Mary was showing true number sentences about the number 9; she was showing things that equal 9." (Each of the examples were explained to the students.) "Now on the bottom half of the page it says 'My number sentences about 8.' When I tell you to start, you are to write number sentences for 8. Work as fast as you can and write as many as you can think of."



Number Sentences

Mary's number sentences about 9.

9 = 7 + 2

9 = 10 - 1

9 = 1 + 5 + 3

My number sentences about 8.

8 = _____ 8 = _____

8 = _____ 8 = _____

8 = _____

8 = _____

8 = _____

8 = _____

8 = _____

8 = _____

8 = _____

	CSMP			Non-CSMP		
	*# True Sentences	# False Sentences	# True Sentences Minus # False Sentences	# True Sentences	# False Sentences	# True Sentences Minus # False Sentences
All Students Combined	8.5	0.9	7.6	6.7	1.4	5.3
Quintile 1	5.2	2.2	3.0	4.4	2.0	2.4
Quintile 2	5.4	0.8	4.6	5.7	2.0	3.7
Quintile 3	7.6	1.0	6.6	5.9	1.2	4.7
Quintile 4	10.7	0.7	10.0	7.8	1.1	6.7
Quintile 5	12.2	0.7	11.5	9.5	0.9	8.6

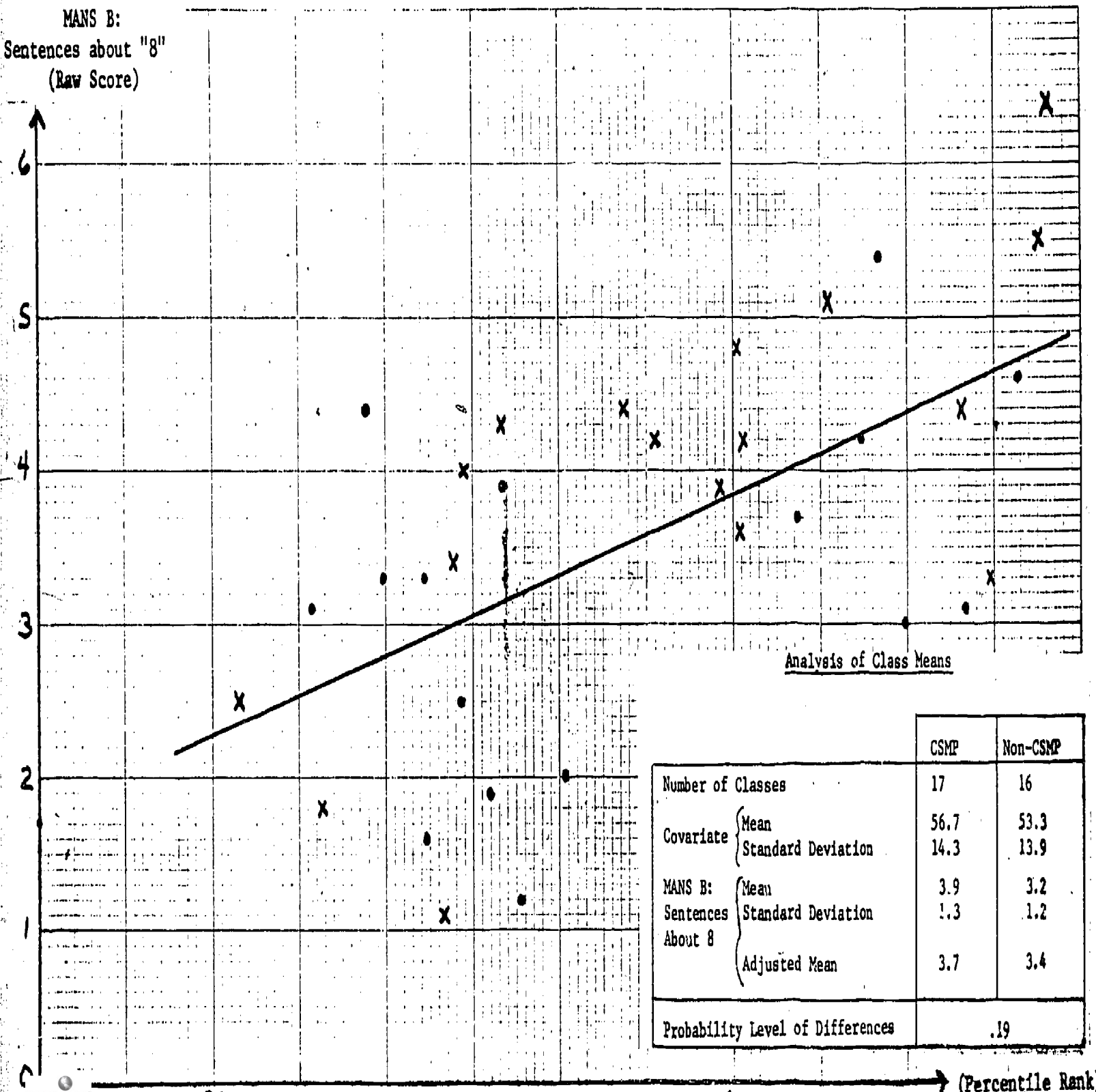
*The graph, next page, used ½x number of true sentences as raw score.

Correlations (Subtest 2 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)		
CSMP-Specific Test A	.66	
CSMP-Specific Test B	.35	
Kuhlmann-Anderson	.65	.58
Reading Tests[1]	.34	.34
Math Tests[1]	.43	.45

[1]Mean correlation from three different standardized reading (math) tests administered at various sites.

MANS Test B: Sentences About "8" versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



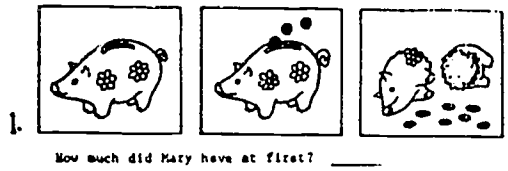
Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		17	16
Covariate	Mean	56.7	53.3
	Standard Deviation	14.3	13.9
MANS B: Sentences About 8	Mean	3.9	3.2
	Standard Deviation	1.3	1.2
	Adjusted Mean	3.7	3.4
Probability Level of Differences		.19	

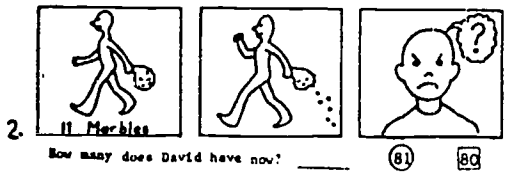
MANS B, Subtest 3: Word Problems

"There are 8 series of pictures on these two pages. For each of three pictures there is a story. At the end of each story there is a question you will be expected to answer. The questions are written directly below each series of pictures. I will read each story aloud two times. Let's all do the first one together." (The students were given time to work on this problem and the answer was explained.)

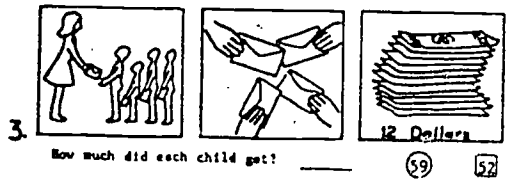
1. "Listen while I tell a story that goes along with the pictures of the piggybank. Look at the first picture. The story goes, 'At first, Mary had some money in her piggybank.' Now look at the second picture. The story continues, 'Mary's father added three cents to her piggybank.' Now look at the third picture. The story goes on, 'Mary broke her piggybank and found seven cents.' Below the pictures it says, 'How much did Mary have at first?' and then there is a blank. I'll tell the story again. At first, Mary had some money in her piggybank. Then Mary's father added 3¢ to her piggybank. Mary broke her piggybank and found 7¢. How much did Mary have at first?"



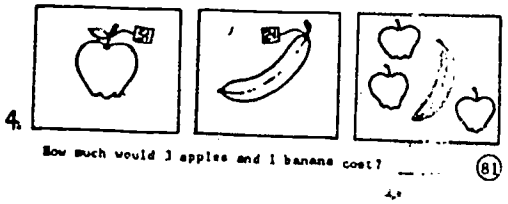
2. "First picture: 'David had 11 marbles in his marble bag.'
Second picture: 'He lost 6 marbles.'
Third picture: 'David is mad at himself.'
Question: 'How many marbles does David have now?"



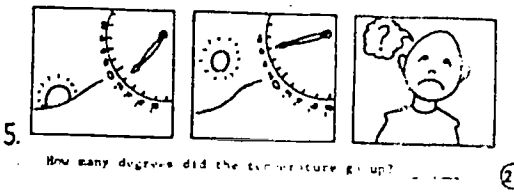
3. "First picture: 'Four children each get the same allowance from their mother.'
Second picture: 'The four children put their allowances together.'
Third picture: 'They have altogether 12 dollars.'
Question: 'How much did each child get?'



4. "First picture: 'One apple costs 5¢.'
Second picture: 'One banana costs 2¢.'
Third picture: 'Sally buys three apples and one banana.'
Question: 'How much would 3 apples and 1 banana cost?'

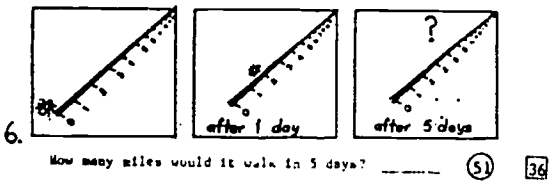


5. "First picture: 'The temperature at sunrise was 2 degrees below zero.'
Second picture: 'The temperature after breakfast was 6 degrees above zero.'
Third picture: 'Amy wonders how much the temperature went up between sunrise and after breakfast.'
Question: 'How many degrees did the temperature go up?'

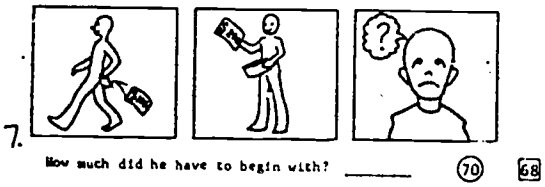


MANS B, Subtest 3: Word Problems (cont.)

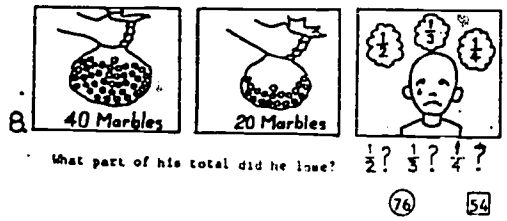
6. "First picture: 'A fantastic ant is about to start a trip.'
Second picture: 'Every day it walks 2 miles.'
Third picture: 'After 5 days it's almost out of sight.'
Question: 'How many miles would it walk in 5 days."



7. "First picture: 'Mr. Rich lost 100 dollars from his wallet.'
Second picture: 'Afterwards he still had 200 dollars.'
Third: 'Mr. Rich wonders how much he had before he lost the 100 dollars.'
Question: 'How much did he have to begin with?"



8. "First picture: 'Donald had 40 marbles before the game.'
Second picture: 'After the game he had only 20 marbles.'
Third picture: 'Donald wonders what fraction of his marbles he lost in the game: $\frac{1}{2}$? $\frac{1}{3}$? $\frac{1}{4}$?'
Question: 'What part of his total did he lose? Did he lose $\frac{1}{2}$ or $\frac{1}{3}$ or $\frac{1}{4}$? You circle the one you think is correct."



MANS B, Subtest 3, Word Problems (cont.)

Means

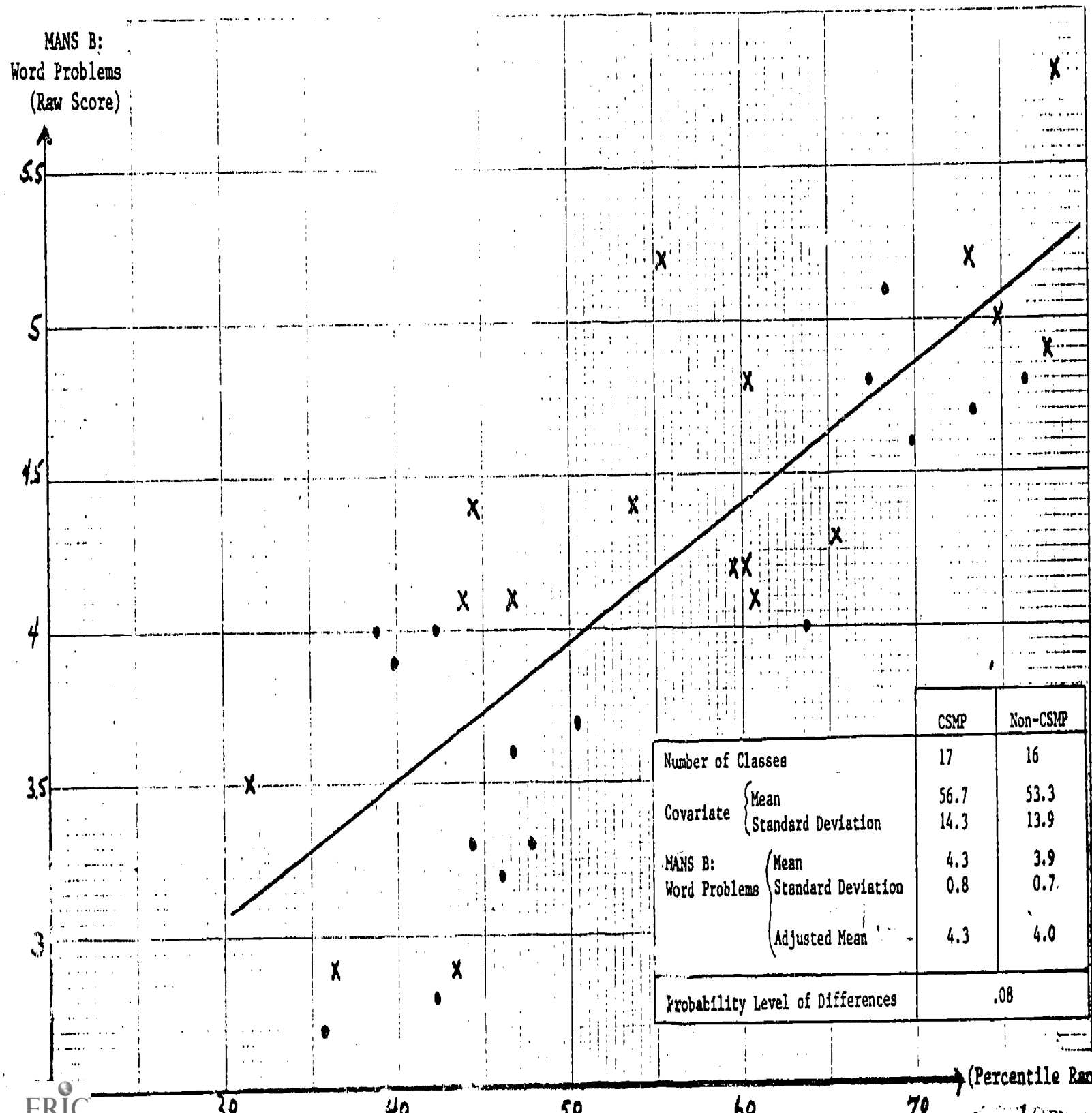
	CSMP	Non-CSMP
All Students Combined	4.4	3.8
Quintile 1	3.1	2.8
Quintile 2	3.5	3.3
Quintile 3	4.0	3.7
Quintile 4	4.8	4.2
Quintile 5	5.4	5.2

Correlations (Subtest 3 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.52	.54
CSMP-Specific Test A	.68	
CSMP-Specific Test B	.22	
Kuhlmann-Anderson	.67	.53
Reading Tests[1]	.37	.35
Math Tests[1]	.45	.53

[1]Mean correlation from three different standardized reading (math) tests administ in various sites.

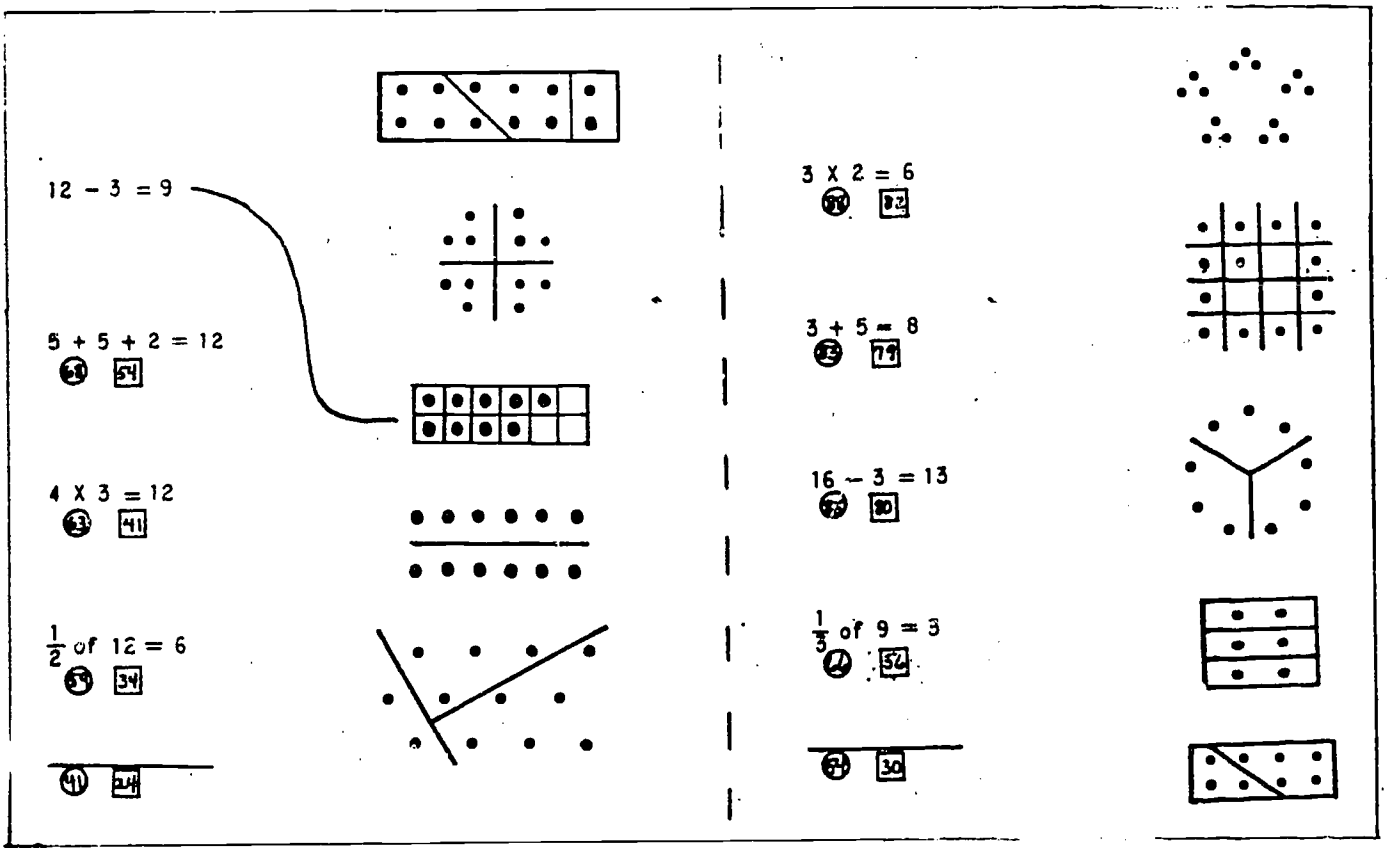
MANS Test B: Word Problems versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



		CSMP	Non-CSMP
Number of Classes		17	16
Covariate	Mean	56.7	53.3
	Standard Deviation	14.3	13.9
MANS B: Word Problems	Mean	4.3	3.9
	Standard Deviation	0.8	0.7
	Adjusted Mean	4.3	4.0
Probability Level of Differences		.08	

MANS B, Subtest 4: Number Sentence Pictures

"On this page there are four number sentences. The first one is 12-3=9, and a line has been drawn from that number sentence to one of the dot pictures on the other side of the page. That is because that dot picture is the best picture on this page for showing that 12-3=9. When I tell you to start, you have to do the same for the next three number sentences. Now after you have done that, there will be one picture left over. You are to make up a number sentence that goes with the picture that's left over, the one that didn't match with any number sentence. Just write your number sentence here on the blank at the bottom of the page. When you have finished this page just go on and do the next page."



Means

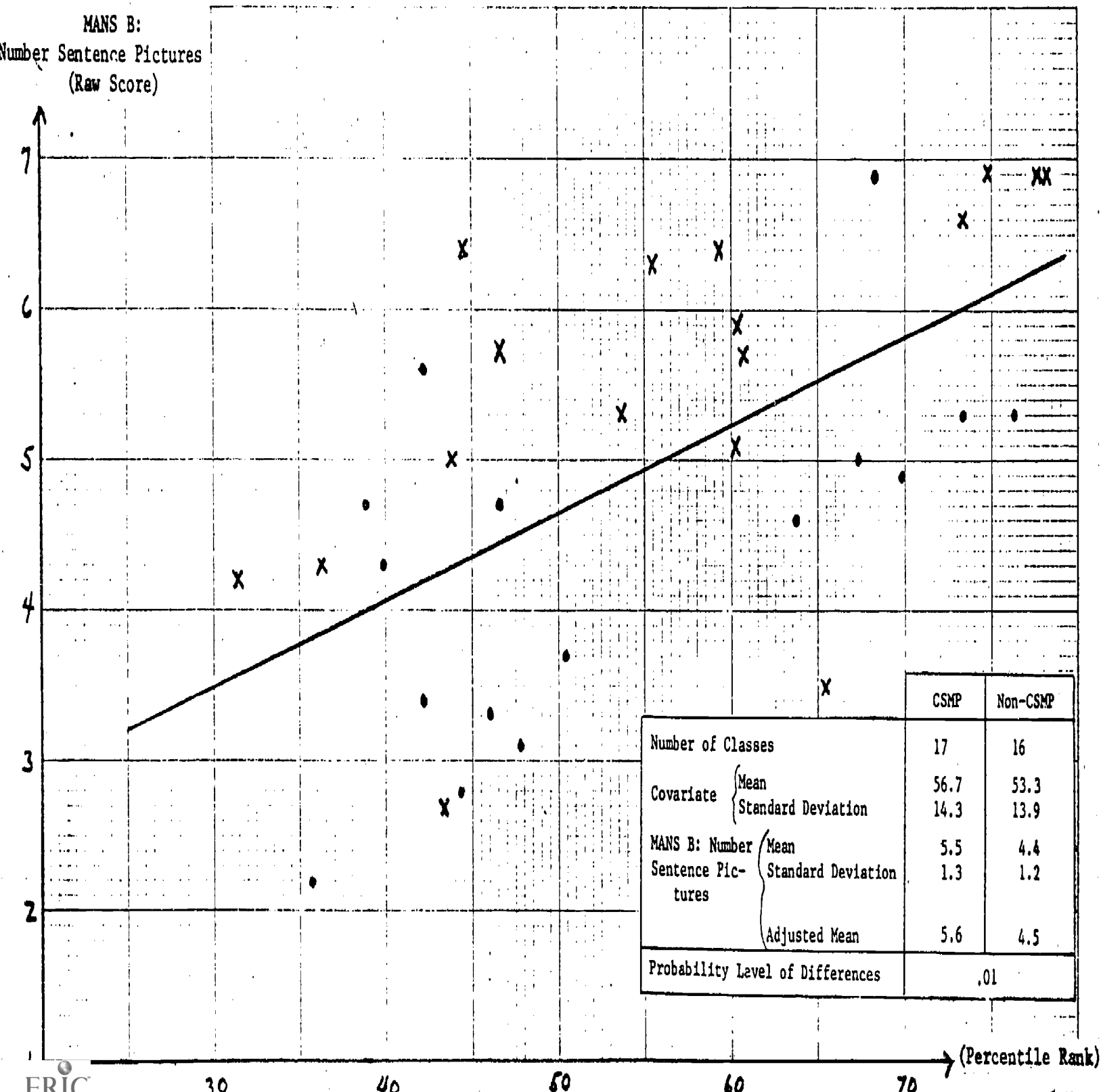
	CSMP	Non-CSMP
All Students Combined	5.7	4.3
Quintile 1	4.4	3.5
Quintile 2	4.2	3.3
Quintile 3	4.7	3.9
Quintile 4	6.3	4.7
Quintile 5	7.3	6.0

Correlations (Subtest 4 v Other Tests)

	CSMP	Non-CSMP
Reliability (KR20)	.79	.70
CSMP-Specific Test A	.69	
CSMP-Specific Test B	.33	
Kuhlmann-Anderson	.53	.52
Reading Tests[1]	.39	.32
Math Tests[1]	.55	.50

[1]Mean correlation from three different standardized reading (math) tests administered at various sites.

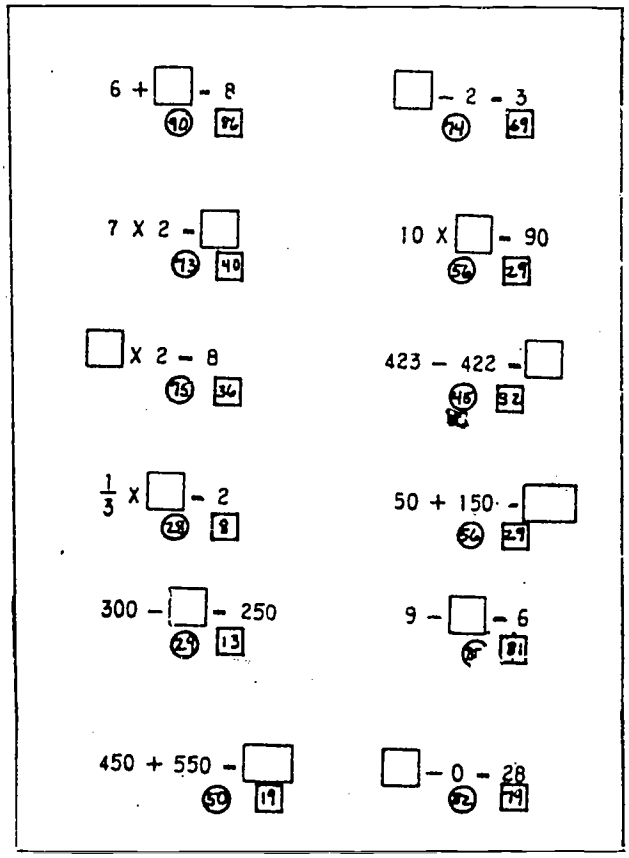
MANS Test B: Number Sentence Pictures versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



		CSMP	Non-CSMP
Number of Classes		17	16
Covariate	Mean	56.7	53.3
	Standard Deviation	14.3	13.9
MANS B: Number Sentence Pictures	Mean	5.5	4.4
	Standard Deviation	1.3	1.2
	Adjusted Mean	5.6	4.5
Probability Level of Differences		.01	

MANS B, Subtest 5: Computation II

"You have to figure out what goes in the boxes to make the number sentences true."



Means

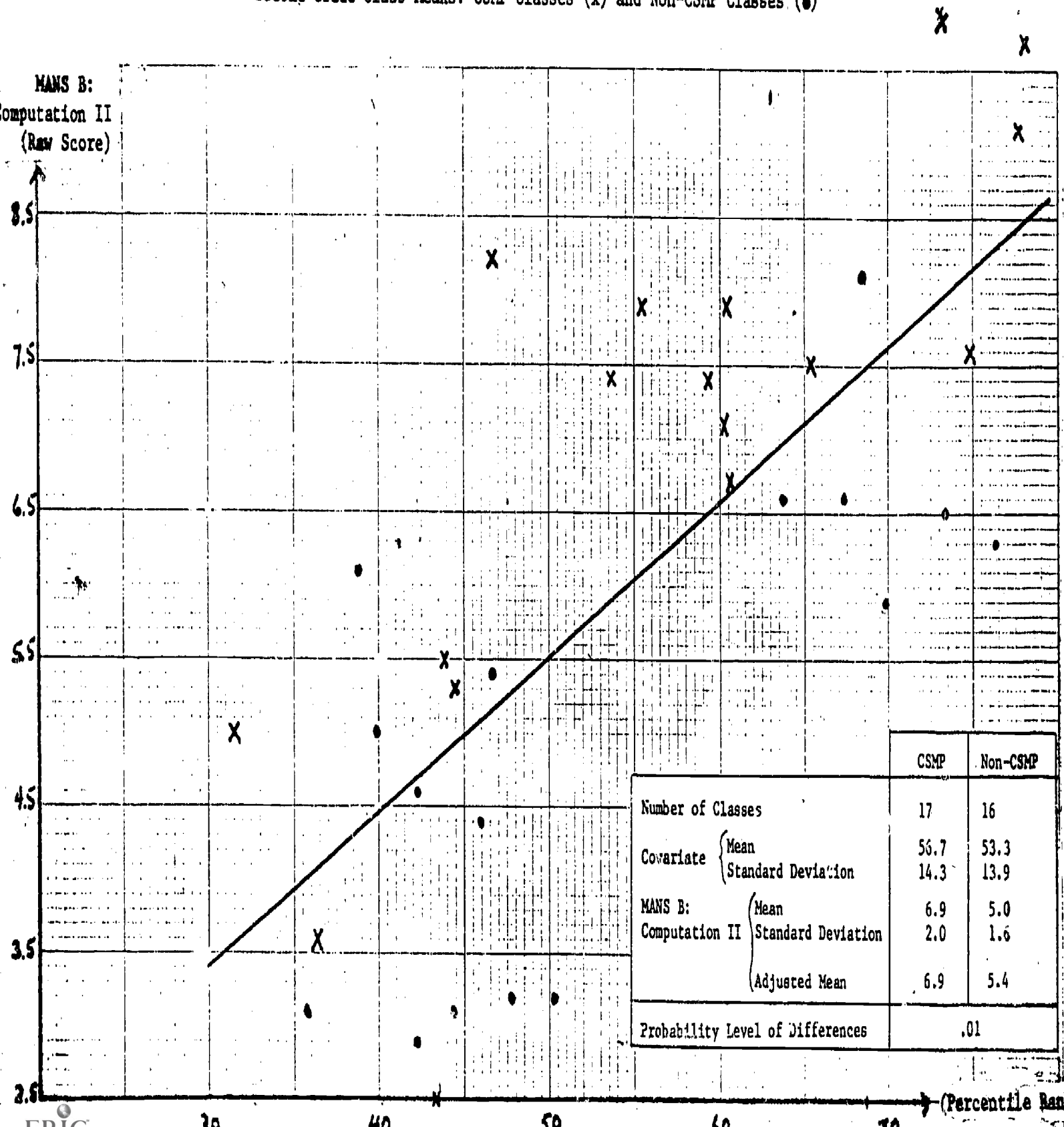
	CSMP	Non-CSMP
All Students Combined	7.1	4.9
Quintile 1	4.3	3.1
Quintile 2	4.7	3.2
Quintile 3	5.9	4.5
Quintile 4	8.3	5.7
Quintile 5	9.4	7.7

Correlations (Subtest 5 v Other Tests)

	CSMP	Non-CSMP
·liability (KR20)	.85	.78
CSMP-Specific Test A	.82	
CSMP-Specific Test B	.53	
Kuhlmann-Anderson	.74	.71
Reading Tests[1]	.39	.49
Math Tests[1]	.67	.71

[1]Mean correlation from three different standardized reading (math) tests administe at various sites.

MANS Test B: Computation II versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



		CSMP	Non-CSMP
Number of Classes		17	16
Covariate	Mean	56.7	53.3
	Standard Deviation	14.3	13.9
MANS B: Computation II	Mean	6.9	5.0
	Standard Deviation	2.0	1.6
	Adjusted Mean	6.9	5.4
Probability Level of Differences		.01	

MANS Test B, Intercorrelations
(First Entry: CSMP, Second Entry: Non-CSMP)[1]

	Subtest 1		Subtest 2		Subtest 3		Subtest 4	
Subtest 1: Labelling Number Lines								
Subtest 2: Sentences About 8	.54	.37						
Subtest 3: Word Problems	.60	.46	.47	.34				
Subtest 4: Number Sentence Pictures	.52	.36	.44	.37	.50	.40		
Subtest 5: Computation II	.70	.60	.56	.44	.67	.56	.56	.55

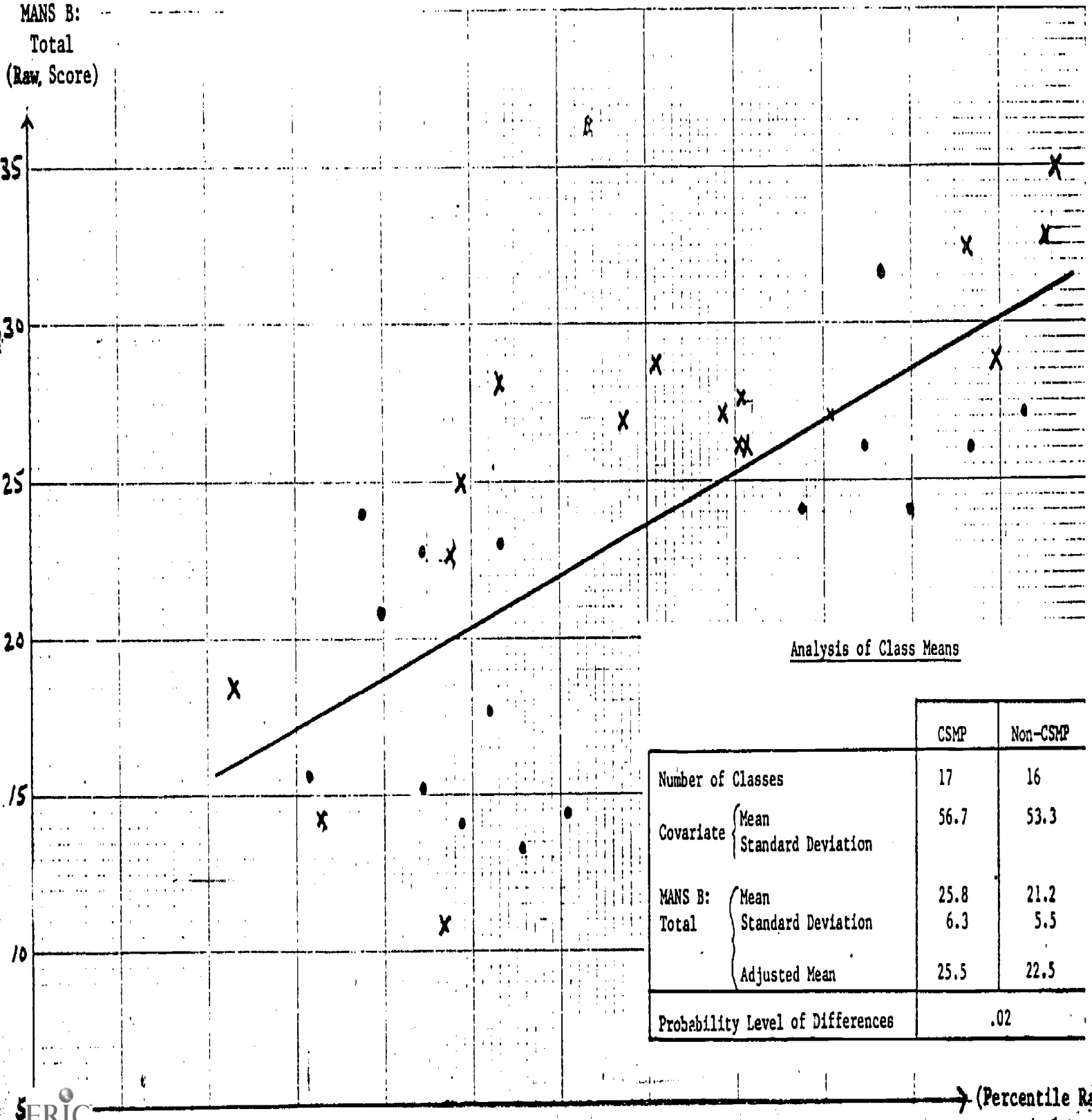
[1]Based on 426 CSMP students and 390 Non-CSMP students.

Correlations Between Total Score, MANS B and Other Tests

	CSMP Students[1]		Non-CSMP Students[1]	
Correlation with: CSMP-Specific Test A	.86	(142)		
CSMP-Specific Test B	.50	(313)		
Kuhlmann-Anderson	.79	(151)	.74	(141)
Gates McGinitie Reading	.53	(61)	.51	(63)
SAT Reading	.71	(97)	.49	(87)
CPT Reading	.27	(47)	.58	(56)
CTBS Math	.80	(250)	.80	(224)
SAT Math	.80	(97)	.78	(87)
CPT Math	.60	(47)	.76	(56)

[1]Number in parenthesis is number of students on whom correlation is based.

MANS Test B: Total Score versus Covariate
Second Grade Class Means: CSMP Classes (x) and Non-CSMP Classes (•)



Analysis of Class Means

		CSMP	Non-CSMP
Number of Classes		17	16
Covariate	Mean	56.7	53.3
	Standard Deviation		
MANS B: Total	Mean	25.8	21.2
	Standard Deviation	6.3	5.5
	Adjusted Mean	25.5	22.5
Probability Level of Differences		.02	

Commentary

Since the two Computations Tests produced significant differences, items from those tests were categorized according to the kind of computation required and are presented below, along with the mean percent correct across categories for CSMP students (first entry) and Non-CSMP students (second entry).

Pure Addition

5+7=☐

50+150=☐

450+550=☐

Percent Correct 63,43

Pure Subtraction

423-422=☐

600-100=☐

9-☐=6

10-☐=7

300-☐=250

Percent Correct 60,54

Pure Multiplication

7x2=☐

2x19=☐

2x400=☐

$\frac{1}{2}$x10=☐

Percent Correct 61,27

Implied Addition

☐-0=28

☐-2=3

☐-49=0

☐-200=100

Percent Correct 64,64

Implied Subtraction

6+☐=8

35+☐=65

Percent Correct 66,67

Implied Division

2x☐=2

10x☐=90

3x☐=300

☐x2=8

☐x5=15

$\frac{1}{3}$x☐=2

Percent Correct 54,31

It can be seen that, although CSMP students scored as high or higher than Non-CSMP students in all categories, the two categories involving multiplication produced the largest differences. Across all items the mean percent correct was 61 (CSMP) and 45 (Non-CSMP). For the 10 items containing a "x" sign, the percentages were 57 and 29; for the remaining items which did not contain a "x" sign the percentages were 64 and 56.

Alternatively, if one looks at the size of the numbers involved, another difference becomes apparent. There were eight items in which at least one of the given numbers exceeded 100. The percentages correct for these items were 51 and 33; for the remaining 16 items the percentages were 66 and 55.

Besides the items in the Computation Tests, there were three items in the MANS tests on which the difference in percent correct in favor CSMP students exceeded 20. These dealt with $\frac{1}{2}$x16 (Test A4), $\frac{1}{2}$ of 12 (B4) and $\frac{1}{2}$ of a bag of marbles (B3). Also, the two items on the Computation Tests which dealt with fractions had percentages correct of 66 versus 13 and

28 versus 8 in favor of CSMP.

Thus, three areas stand out in which CSMP students did far better than Non-CSMP students: fractions (or at least calculating a fractional part of), multiplication and working with numbers larger than 100. These areas reflect the larger-than-usual emphasis which is apparent in the CSMP curriculum materials. There were of course many other test items which produced fairly large differences, but such differences were not so dramatic.